Psychological Review

CHARLES N. COFER, EDITOR
Pennsylvania State University

VOLUME 73, 1966

HELEN ORR
Managing Editor

ELIZABETH MULLEN
Editorial Assistant

HENRY S. SALMON
Advertising and Business Manager

ROBERT D. GORMAN
Subscription Manager

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA. 17604
AND 1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

Editorial Consultants

Jack A. Adams
Robert E. Adamson
Norman H. Anderson
Fred Attneave
George Barlow
William S. Battersby
Frank A. Beach
D. E. Berlyne
Roger W. Black
Donald Blough
William H. Bogartz
Lyle E. Bourne
Weston A. Bousfield
Gordon H. Bower
Jack W. Brehm
Judson S. Brown
Arnold H. Buss
E. C. Carterette
Dorwin Cartwright
Alphonse Chapanis
Loren J. Chapman
Burton H. Cohen
Jacob Cohen
M. R. D'Amato
James Deese
William N. Dember
Victor H. Denenberg
Joel Elkes
William Epstein
Charles W. Eriksen
Donald W. Fiske
John H. Flavell
Edwin A. Fleishman
Leo Ganz
Eleanor I. Gibson
Robert Glaser
Richard C. Gonzalez
Jacqueline I. Goodnow
Isidore Gormezano
Albert E. Goss
David M. Green
G. Robert Grice
Norman Guttman
John F. Hall
Kenneth R. Hammond
Harry F. Harlow
Keith J. Hayes
Robert C. Haygood
D. O. Hebb
R. J. Herrnstein
Roger W. Heyns
Wayne H. Holtzman
Fred E. Horvath
J. McV. Hunt
Leo M. Hurvich
Murray E. Jarvik
James J. Jenkins
Harry I. Kalish
Lawrence Karlin
Geoffrey Keppel
William Kessen
Gregory A. Kimble
John I. Lacey
Roy Lachman
Herschel W. Leibowitz
Seymour Levine
Donald J. Lewis
Donald B. Lindsley
Frank A. Logan
R. Duncan Luce
Robert B. MacLeod
Robert B. Malmo
Edwin Martin
Melvin H. Marx
F. Joseph McGuigan
Arnold Mechanic
Arthur W. Melton
Robert E. Miller
Harold E. Mitzel
Jane S. Mouton
Conrad J. Mueller
Bennet B. Murdock, Jr.
Jerome L. Myers
J. Nachmias
Theodore M. Newcomb
Kenneth Neil Ogle
David S. Palermo
Robert W. Payne
Carl Pfaffman
Howard R. Pollio
William F. Prokasy
Allan Rechtschaffen
Frank Restle
Harriet L. Rheingold
Benbow F. Ritchie
Julian B. Rotter
Rudolph W. Schulz
Robert Seibel
John P. Seward
Seth K. Sharpless
Roger N. Shepard
Jerome E. Singer
Norman J. Slamecka
Richard L. Solomon
Kenneth W. Spence
Charles D. Spielberger
Arthur W. Staats
Eliot Stellar
John Theios
John W. Thibaut
Richard F. Thompson
Warren S. Torgerson
Thomas Trabasso
Endel Tulving
Allan R. Wagner
Edward L. Walker
Nancy C. Waugh
Wilse B. Webb
Sidney Weinstein
Robert W. White
Ruth C. Wylie
Richard P. Youtz
Philip G. Zimbardo
Joseph Zubin

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 73

Baron, Reuben M. Social Reinforcement Effects as a Function of Social Reinforcement History 527
Beckwith, Mary, and Restle, Frank. Process of Enumeration 437
Brelsford, John W., Jr. See Theios, John.
Brody, Nathan, and Oppenheim, Paul. Tensions in Psychology between the Methods of Behaviorism and Phenomenology 295
Broen, William E., Jr. Response Disorganization and Breadth of Observation in Schizophrenia 579
Broen, William E., Jr., and Storms, Lowell H. Lawful Disorganizaton: The Process Underlying a Schizophrenic Syndrome 265
Broen, William E., Jr. See Storms, Lowell H.
Cairns, Robert B. Attachment Behavior of Mammals 409
Campbell, Byron A., and Jaynes, Julian. Reinstatement 478
Capaldi, E. J. Partial Reinforcement: A Hypothesis of Sequential Effects 459
Carrington, Patricia. See Ephron, Harmon S.
Cohen, Bertram D. See Rosenberg, Seymour.
Day, Ruth S. See Horowitz, Leonard M.
DeNike, L. Douglas. See Spielberger, Charles D.
Ekstrand, Bruce R., Wallace, William P., and Underwood, Benton J. A Frequency Theory of Verbal-Discrimination Learning 566
Ekstrand, Bruce R. See Underwood, Benton J.
Ephron, Harmon S., and Carrington, Patricia. Rapid Eye Movement Sleep and Cortical Homeostasis 500
Freeman, Robert B., Jr. Optical Texture versus Retinal Perspective: A Reply to Flock 365
Games, Paul A. Comments on "A Power Comparison of the *F* and *L* Tests—I" .. 372
Ganz, Leo. Mechanism of the Figural Aftereffects 128
Haber, Ralph Norman. Nature of the Effect of Set on Perception 335
Hakes, David T., and Young, Robert K. On Remote Associations and the Interpretation of Derived-List Experiments 248
Hay, John C. Optical Motions and Space Perception: An Extension of Gibson's Analysis 550
Holman, Eric W. Tests for Spontaneous Alternation 427
Horowitz, Leonard M., Norman, Sandra A., and Day, Ruth S. Availability and Associative Symmetry 1
Irwin, Francis W. Criteria of Expectancy 327
Irwin, Francis W. Reply to Paul Thomas Young 241
Jakobovits, Leon A. Mediation Theory and the "Single-Stage" S–R Model: Different? 376
Jaynes, Julian. See Campbell, Byron A.
Kessen, William. See Munsinger, Harry.
Kimmel, H. D. Inhibition of the Unconditioned Response in Classical Conditioning 232
Lewis, Donald J., and Maher, Brendan A. Electroconvulsive Shock and Inhibition: Some Problems Considered 388
Lieberman, Lewis R. On "Actuarial Methods as Appropriate Strategy for the Validation of Diagnostic Tests" 262
Lovejoy, Elijah. Analysis of the Overlearning Reversal Effect 87
Maher, Brendan A. See Lewis, Donald J.
Makous, W. L. Cutaneous Color Sensitivity: Explanation and Demonstration 280
McGaugh, James L., and Petrinovich, Lewis L. Neural Consolidation and Electroconvulsive Shock Reexamined 382

McKee, John P. See Riley, Donald A.
Munsinger, Harry, and Kessen, William. Stimulus Variability and Cognitive Change 164
Norman, Sandra A. See Horowitz, Leonard M.
Oppenheim, Paul. See Brody, Nathan.
Parks, Theodore E. Signal-Detectability Theory of Recognition-Memory Performance 44
Peterson, Lloyd R. Short-Term Verbal Memory and Learning 193
Petrinovich, Lewis L. See McGaugh, James L.
Piaget, Jean. Response to Brian Sutton-Smith 111
Pinneo, Lawrence R. On Noise in the Nervous System 242
Restle, Frank. See Beckwith, Mary.
Riley, Donald A., Sherman, Marian, and McKee, John P. A Comment on Intermediate Size Discrimination and Adaptation-Level Theory 252
Rosenberg, Seymour, and Cohen, Bertram D. Referential Processes of Speakers and Listeners 208
Routtenberg, Aryeh. Neural Mechanisms of Sleep: Changing View of Reticular Formation Function 481
Sherman, Marian. See Riley, Donald A.
Spielberger, Charles D., and DeNike, L. Douglas. Descriptive Behaviorism versus Cognitive Theory in Verbal Operant Conditioning 306
Spence, Kenneth W. Cognitive and Drive Factors in the Extinction of the Conditioned Eye Blink in Human Subjects 445
Spencer, William Alden. See Thompson, Richard F.
Stein, Larry. Habituation and Stimulus Novelty: A Model Based on Classical Conditioning 352
Storms, Lowell H., and Broen, William E., Jr. Drive Theories and Stimulus Generalization 113
Storms, Lowell H. See Broen, William E., Jr.
Sutton-Smith, Brian. Piaget on Play: A Critique 104
Theios, John, and Brelsford, John W., Jr. A Markov Model for Classical Conditioning: Application to Eye-Blink Conditioning in Rabbits 393
Thompson, Charles P. On the Incompatibility of the Houston and Osgood Transfer Surfaces 586
Thompson, Richard F., and Spencer, William Alden. Habituation: A Model Phenomenon for the Study of Neuronal Substrates of Behavior 16
Underwood, Benton J., and Ekstrand, Bruce R. An Analysis of Some Shortcomings in the Interference Theory of Forgetting 540
Underwood, Benton J. See Ekstrand, Bruce R.
Wallace, William P. See Ekstrand, Bruce R.
Weitzman, R. A. Statistical Learning Models and Individual Differences 357
Whalen, Richard E. Sexual Motivation 151
Wheeler, Ladd. Toward a Theory of Behavioral Contagion 179
Young, Paul Thomas. Hedonic Organization and Regulation of Behavior 59
Young, Robert K. See Hakes, David T.
Zeiler, Michael D. The Stimulus in the Intermediate Size Problem 257

VOL. 73, NO. 1 JANUARY 1966

PSYCHOLOGICAL REVIEW

AVAILABILITY AND ASSOCIATIVE SYMMETRY[1]

LEONARD M. HOROWITZ, SANDRA A. NORMAN, AND RUTH S. DAY

Stanford University

This paper examines the concept of availability and its theoretical role in the cognitive processes. An item's availability (AV) is operationally defined by the probability that S could recall it after a 15-sec. delay. (AV seems to grow fastest when S produces it from memory. It also grows, though not as fast, when S sees the item without producing it.) This definition is used to examine the principle of associative symmetry, and evidence is presented to support it. Sources of asymmetry in natural language are examined, and the concept's theoretical implications for memory and thought are discussed.

According to linguists, a word can be defined by its boundaries in a sentence. For Hockett (1958) a word is "bounded by successive points *at which pausing is possible* [p. 167]." Pauses never occur in the middle of a word; a pause would destroy the gestalt of the word, and a word's gestalt distinguishes it from a mere collection of speech sounds. Thus, a brand new word might be analyzed into speech sounds, but after repetitions the word as a whole becomes psychologically primary. Its properties no longer depend on properties of the ingredient sounds. This shift from the sound as a unit to the whole word as a unit is a remarkable fact of language, a fact which may help explain the versatility of memory and thought. The present authors assume that this shift is gradual and continuous. And the word is said to grow more and more "available."

The term "availability" has been used in two different senses. Perhaps in this paper it should be called "item availability" because it is used to describe a single verbal unit. Two recent authors have also used availability in this way. According to Underwood, Runquist, and Schulz (1959), an available item is "readily recallable"; for Asch and Ebenholtz (1962b) it is "accessible to recall." Item availability denotes how readily a verbal unit comes to mind as a label for experience, how salient it is as a unit of speech, or how often it occurs in verbal behavior.

Availability, in the second sense, denotes an associate's position in some response hierarchy. The higher the associate in the hierarchy, the more available it is said to be. Perhaps "hierarchy availability" would better convey this meaning. Hierarchy avail-

---

[1] This research was supported in part by United States Public Health Service Grant HD 00922-03.

TABLE 1

INITIAL MEASURES OF AVAILABILITY

Test item	Value of AV
ZUMAP	.00
LATUK	.04
SAGROLE	.04
NEGLAN	.08
TAROP	.08
QUIPSON	.16
GOJEY	.20
POLEF	.24
VOLVAP	.24
MEARDON	.28

ability resembles the concept of response strength, but the terms have slightly different meanings. Response strength usually describes the reinforcement history of an S-R bond. An item's response strength does not change abruptly from time to time. Hierarchy availability, on the other hand, connotes variability in an item's hierarchy position. Under different stimulating conditions, different responses ascend. As one example, the hierarchy changes drastically if the stimulus is placed in a different context (Howes & Osgood, 1954). Or second, a given response in the hierarchy can be "primed": If some irrelevant stimulus elicits a response, that response becomes more probable later in free association. Also, as this article shows, a response typically ascends in a hierarchy whenever its *item* availability increases.

Item availability is chosen as the more fundamental concept since it is independent of any particular response hierarchy. In general, an item grows more available as the subject has more experience with it. But the critical growth variables have not been known, so the concept has not been defined precisely. This article first proposes an operational definition of item availability, then shows how an item's availability can be increased empirically, and, finally, examines the theoretical role of availability in the cognitive processes.

AN OPERATIONAL DEFINITION

A "test item" is defined as a collection of letters whose availability was to be measured. Test items in the present studies were always two-syllable nonsense words. Each was printed on a 4 × 6-inch index card as the middle of three items. The items were arranged vertically, and the two extra ones comprised the "test frame." Thus, the test item NEGLAN was embedded in a test frame to form the triad BALAP NEGLAN GOKEM. All three items appeared in large capital letters, and the subject studied the display for 2 seconds.

Then the card was reversed to expose a three-digit number. The subject repeated this number aloud and then started counting backwards by 3's. A metronome was set for 120 beats per minute, and the subject counted in rhythm with every other beat. After 15 seconds, the subject stopped counting and tried to recall the triad of items. Twenty-five subjects were tested in this way. The test item's availability (AV) was then defined as the proportion of the subjects who recalled it correctly after the delay. This procedure is based on the experimental method of Peterson and Peterson (1959). It follows Hunter (1913, pp. 52–62) in trying to measure how well the nervous system sustains symbolic activity over time.

Ten test items were measured in this way. They were all selected from Noble's (1952) list to have *m* values of 2.00 or less. Each test item was systematically embedded in 10 different test frames. A test item was assigned to a different test frame from one subject to the next, and the order of presentation was systematically varied. Resulting AV values are presented in Table 1.

HOW TO INCREASE AVAILABILITY

The following experiment (Experiment I) was designed to increase the above values. The results show that an item's AV grows fastest if the subject produces the item from memory. Availability also grows, though not as rapidly, when the subject sees the item without producing it.

First, the subject learned a PA task with four pairs: A-B, B-C, D-E, and F-E. In this task, B served a dual role; it was the stimulus of one pair and the response of another. Item E served two roles as well, the response of two pairs. The following diagram describes the associative relationship among items. An arrow in the diagram leads from a PA stimulus to a PA response.

$$\begin{tikzcd} & & & & F \arrow[d] \\ A \arrow[r] & B \arrow[r] & C & D \arrow[r] & E \end{tikzcd}$$

The vocabulary of this task was selected from Table 1. For one group of subjects (Experiment Ia) the following nonsense words were adopted: LATUK, A; QUIPSON, B; VOLVAP, C; TAROP, D; POLEF, E; and MEARDON, F. For another group of subjects (Experiment Ib) the vocabulary was: VOLVAP, A; GOJEY, B; LATUK, C; POLEF, D; NEGLAN, E; and SAGROLE, F. Words were assigned to the two PA tasks so as to counterbalance initial item availabilities. Thus, the AV of Item A was .04 in Experiment Ia and .24 in Experiment Ib, averaging .14. Similar averages for other ingredients of the associative structure ranged from .14–.18.

The four pairs were presented on a Lafayette memory drum. First, each pair of associates coappeared for 4 seconds. On each later trial the subject tried to anticipate the response. The stimulus appeared for 2 seconds, and the response then joined it for 2 more seconds. The interval between pairs lasted 4 seconds; the interval between trials, 12 seconds. The pairs were presented in four different orders, systematically varied from trial to trial. The subject tried to anticipate the response before it appeared, and the procedure lasted 10 anticipation trials. Twenty-five subjects were tested in each subexperiment. At least one perfect trial occurred somewhere in the performance of 17 subjects in Experiment Ia and 16 subjects in Experiment Ib.

After the PA task, each item's availability was measured by the procedure described above. Four control items which the subject had not seen before were added to make a batch of 10 test items. In Experiment Ia, the control items were GOJEY, SAGROLE, NEGLAN, and ZUMAP; in Experiment Ib, TAROP, QUIPSON, ZUMAP, and MEARDON.

The resulting AV values for all items are presented in Table 2. The highest value was only .84, so the items probably did not reach the level of bonafide words. Only one critical variable seemed to determine an item's AV level—namely, its role in the PA task. "Produced" items are items which the subject had produced from memory, "seen" items had just been stimuli, and "control" items had not appeared in the PA task at all. Averaged across the two experiments, produced items earned a mean AV value of .72; seen items, .50; control items, .18. Within each experiment the difference between seen and produced items in

TABLE 2

Measures of Availability after PA Task

PA role	Experiment Ia			Experiment Ib		
	Item	Value of AV	M	Item	Value of AV	M
Produced	QUIPSON (B)	.84	.75	GOJEY(B)	.76	.68
	VOLVAP (C)	.76		LATUK (C)	.72	
	POLEF (E)	.64		NEGLAN (E)	.56	
Seen	LATUK(A)	.52	.56	VOLVAP (A)	.56	.44
	TAROP(D)	.40		POLEF (D)	.32	
	MEARDON (F)	.76		SAGROLE (F)	.44	
Control	GOJEY	.12	.11	MEARDON	.48	.24
	NEGLAN	.12		QUIPSON	.24	
	SAGOLE	.08		TAROP	.20	
	ZUMAP	.12		ZUMAP	.04	

mean AV was significant. In Experiment Ia, $z = 2.44$, $p < .02$; in Experiment Ib, $z = 2.96$, $p < .01$. Likewise, the difference between seen and control items was significant in each experiment. In Experiment Ia, $z = 6.41$, $p < .001$; in Experiment Ib, $z = 2.79$, $p < .01$. Item E, the response to two different stimuli, seems to have earned a slightly lower AV value than the other two produced items. This surprising result appeared in both experiments and needs to be examined further.

Thus, in general, an item's availability increases when the subject produces it from memory. Articulating an item is no more effective than looking at it (Underwood, 1964), so the subject must have to produce the item *from memory* to raise its AV. Producing the item from memory seems to unify the ingredient sounds into a larger gestalt. And this growth in AV does not depend on the item's associates. Thus, free, serial, or PA recalls would all heighten item availabilities. PA learning is especially interesting since the subject only produces the *response* from memory, and the response thus becomes more available than the stimulus. PA stimuli grow more available, too, but according to Experiment I, not as greatly.

Principle of Associative Symmetry

Availability deserves to be measured because of a recent issue in the study of cognition. Asch and Ebenholtz (1962b) and Horowitz, Brown, and Weissbluth (1964) have shown that the verbal association process may differ fundamentally from classical conditioning. While a conditioned response is directional—a stimulus elicits a response—a verbal association may be symmetrical. According to this hypothesis, an association between two verbal units, A-B, is symmetrical if A and B are equally available. Equally available associates are supposed to elicit each other with equal probabilities. Thus, an association may be viewed as a pair of equipotent S-R bonds or perhaps as a union of two elements into a new gestalt. The authors find either view tenable.

The principle further asserts that an association exhibits direction if the associates differ in availability. If so, the less-available associate seems to elicit the more-available one, but not vice versa. Phenomenally, then, the association would resemble a conditioned response. In PA learning, especially, the associates differ in availability, and the less-available stimulus would seem to elicit the more-available response. According to the principle, however, the latent backward association could be made overt by any operation that increased the stimulus availability.

To test the principle of associative symmetry, Horowitz, Brown, and Weissbluth (1964) prepared a PA task that contained six pairs of nonsense words. The stimuli of some pairs were responses of other pairs, so these stimuli became highly available. (One pair, for example, was denoted C-E; its stimulus served as a response in the pair A-C and thus became highly available.) After the subject mastered the PA task, he was asked for two free associations to each ingredient of the task's vocabulary. Response hierarchies were prepared to show how often every item elicited every other item as associate.

In every hierarchy the strongest responses were available ones. The most popular associate any item elicited was an available one which had coappeared with it during learning. The "direction" of pairing, though, was unimportant. The association C-E, for example, was supposed to be symmetrical. Therefore, the probability that C elicited E in free association was compared with the probability that E elicited C. These probabilities did not differ significantly: The former was .41, the latter, .39. The subject had also learned a pair, E-G. The probability that E elicited G was compared with the probability that E elicited C. These figures did not differ significantly either: The former equaled .42, the latter, .39. Comparisons like these strongly supported the principle of associative symmetry.

When item availabilities did differ, the data revealed marked directionalities. The stimulus of A-C was less available since A was not produced. The probability that A elicited C in free association was therefore compared with the probability that C elicited A. Whereas the former probability equaled .49, the latter was only .12.

A REFINED REPLICATION

These comparisons were all comparisons within subjects. Horowitz, Brown, and Weissbluth (1964) did not operate on a single association to vary just its directionality in different subjects. The following experiment (Experiment II) achieved this goal and also replicated the above results. This experiment compared the free associations of two experimental groups whose learning tasks differed slightly. Their associative structures are diagramed below to describe each group's PA task. Again, an arrow in the diagram leads from a PA stimulus to its PA response. Each task contained five pairs of associates with four items serving as responses.

Condition 1

$$\begin{tikzcd}
D \arrow[d] & E \arrow[d] & & \\
A \arrow[r] & B \arrow[r] & C \arrow[r] & F
\end{tikzcd}$$

Condition 2

$$\begin{tikzcd}
 & E \arrow[d] & & & & \\
A \arrow[r] & B \arrow[r] & C \arrow[r] & F & D \arrow[r] & X
\end{tikzcd}$$

The major difference focused on Item A, which became more available in Condition 1. From the principle of associative symmetry, the backward association B-A should only have appeared in Condition 1. In Condition 1 alone its free-association incidence was to equal the incidence of A-B and B-C. B-C and C-F, though, should have been symmetrical in both experimental conditions.

An association was also expected to develop between Items A and E. These items were both stimuli for B during learning. Kjeldergaard and Horton's (1960) "stimulus equivalence" paradigms would predict such an association, and so would the principle of induction described below. Condition 2 did not make either associate available, though, so both directions of this association were expected to stay latent. Condition 1, however, allowed E to elicit the more available A and thus reveal a directional association.

Twenty-five subjects were assigned to each condition. First each subject learned a PA task with nonsense words as the vocabulary (Set N). The nonsense words were selected from Noble's (1952) list to have *m* values of 2.00 or less. They were: VOLVAP, A; TAROP, B; QUIPSON, C; MEARDON, D; NEGLAN, E; GOJEY, F; and LATUK, X.

After the subject mastered the learning task, he received a booklet for his free associations. Each page contained one item of the task, and this item appeared on two consecutive pages. The subject was instructed to supply two associations to every item. A hierarchy of associations was then compiled for each item to show all of the associations and their probabilities of occurrence.

Then the subject learned another PA task based on the same associative structure. This task's vocabulary was composed of low-frequency English words (Set W). These words were also selected from Noble's (1952) list to have *m* values of 2.00 or less. They were: TUMBRIL, A; XYLEM, B; RENNET, C; ATTAR, D; BODKIN, E; WIDGEON, F; and DAVIT, X. After the subject mastered this PA task, another word-association test was administered, and response hierarchies for Set W were compiled.

Pairs of associates were presented on a Lafayette memory drum. The stimulus was exposed alone for 2 seconds, and then the stimulus and response coappeared for 2 seconds. A 4-second interval separated pairs, and the intertrial interval lasted 12 seconds. Four different orders of presentation were adopted, and the task continued to a criterion of two successive perfect trials.

The principal data concern the subject's responses on the association test. Each subject produced two associations per item in each set. An item's *response hierarchy* is then defined as its frequency distribution of responses. A response hierarchy contained 50 responses; for convenience the response frequencies are reported as proportions. The value of r between corresponding proportions of Set N and Set W was .89 for Condition 1 and .90 for Condition 2. Because of the strong agreement between sets, corresponding values were averaged.

The stimulus of each hierarchy is called the hierarchy's *title* to avoid theoretical misunderstandings. Hereafter, each experimental result is labeled by a pair of letters: The first letter denotes the hierarchy's title, and the second, one of its associates. Thus, for example, the probability of AC refers to the proportion of times A elicited C on the association test. Table 3 reports all of these probabilities.

Backward associations were popular when the associates were both highly available. BA, for example,

TABLE 3

Response Hierarchies from Free-Association Test

Condition 1

Title	Responses	Proportionate frequency of occurrence		
		Set N	Set W	M
A	B	.40	.40	.40
	C	.30	.16	.23
	D	.14	.28	.21
	F	.10	.14	.12
	E	.06	.02	.04
B	C	.34	.38	.36
	A	.32	.32	.32
	F	.20	.12	.16
	D	.08	.08	.08
	E	.06	.10	.08
C	F	.36	.40	.38
	B	.32	.34	.33
	A	.22	.10	.16
	D	.08	.06	.07
	E	.02	.10	.06
D	A	.44	.54	.49
	B	.28	.32	.30
	F	.10	.10	.10
	C	.16	.02	.09
	E	.02	.02	.02
E	B	.46	.48	.47
	A	.28	.22	.25
	F	.14	.14	.14
	C	.08	.12	.10
	D	.04	.04	.04
F	C	.50	.44	.47
	B	.32	.20	.26
	A	.06	.18	.12
	E	.08	.08	.08
	D	.04	.10	.07

Condition 2

Title	Responses	Proportionate frequency of occurrence		
		Set N	Set W	M
A	B	.42	.52	.47
	C	.24	.16	.20
	X	.20	.08	.14
	F	.06	.10	.08
	D	.06	.10	.08
	E	.02	.04	.03
B	C	.38	.40	.39
	E	.14	.30	.22
	A	.12	.12	.12
	F	.12	.08	.10
	X	.16	.04	.10
	D	.08	.06	.07
C	F	.40	.40	.40
	B	.36	.42	.39
	X	.14	.06	.10
	E	.06	.08	.07
	A	.02	.04	.03
	D	.02	.00	.01
D	X	.50	.52	.51
	B	.24	.20	.22
	F	.10	.10	.10
	C	.08	.10	.09
	E	.04	.04	.04
	A	.02	.04	.03
E	B	.50	.44	.47
	F	.12	.20	.16
	C	.08	.18	.13
	X	.18	.08	.13
	D	.08	.06	.07
	A	.04	.04	.04
F	C	.48	.42	.45
	B	.20	.35	.28
	X	.24	.08	.16
	E	.02	.10	.06
	D	.06	.00	.03
	A	.00	.04	.02
X	D	.34	.38	.36
	B	.26	.30	.28
	C	.16	.14	.15
	F	.10	.12	.11
	E	.10	.02	.06
	A	.04	.04	.04

was very probable in Condition 1 where A was highly available; its probability, averaged across sets, was .32. This figure did not differ significantly from the probability of BC (.36), nor from the probability of AB (.40). The former $z = .67$, the latter, 1.18; both p's $> .20$. In Condition 2, though, where A was less available, the probability of BA was .12, a significant reduction from its value in Condition 1; $z = 3.41$, $p < .001$.

In both experimental conditions Items B, C, and F were highly available. Therefore, the incidences of BC and CF were expected to equal those of CB and FC, the backward associations. The probability of BC was .36 in Condition 1, in Condition 2, .39. The corresponding probabilities of CB were .33 and .39. Similarly, the probability of CF was .38 in Condition 1 and .40 in Condition 2. The corresponding probabilities of FC were .47 and .45. No pair of differences here was statistically significant, all z's $<$ 1.29, $p > .10$. The slightly greater incidence of FC (a backward association) may be due to F's unique position in the associative structure (Horowitz, Brown, & Weissbluth, 1964).

An *induced association* is defined as an association that develops between two items which never coappeared during learning. Since this term is neutral theoretically, it is preferred to "mediation." The principle of induction asserts that an association becomes induced between two items which have shared a common linguistic context (Horowitz, Brown, & Weissbluth, 1964). Of course, the induced association only becomes overt when one associate or both are highly available. This principle explains how paradigmatic associations develop (McNeill, 1963). It tells why a high-frequency adjective becomes associated with its polar opposite, even though the two words do not usually coappear in the language.

An induced association was expected between Items E and A. This association was only to be evident when A was available. And, indeed, the probability of EA was .25 in Condition 1 but only .04 in Condition 2. The difference was statistically significant, $z = 4.22$, $p < .001$. Furthermore—and this is important—EA was more probable than EC in Condition 1, $z = 2.78$, $p < .01$. While EA had satisfied the conditions for induction, EC had not.

RESPONSE LATENCIES

Asch and Lindner (1963) have implied that a response's latency depends on its availability. If the two associates are equally available, their two directions of recall should have equal latencies. A subject is said to perform a "forward recall" of an association, A-B, when he observes A and supplies B. He performs a "backward recall" when he observes B and supplies A. If A and B are equally available, the forward and backward recall latencies should be equal. Experiment III was designed to examine this hypothesis.

The associative structure of Condition 1 above was adopted with Set W's vocabulary. Twenty-six subjects learned this PA task to a criterion of one perfect trial; the procedure exactly replicated the one described above. Then each subject was tested for forward and backward recall. He was given a card that contained one word of Set W and a blank space next to it. Sometimes the blank appeared to the word's left, sometimes to its right. The subject had to say which word belonged in the blank. Thus, every stimulus and every response was presented alone once, and the subject had to supply its associate. The card with Item B in the response position had two correct answers because of the two learned pairs, A-B and E-B. When this card appeared, the subject was asked to give both correct answers.

Each subject could produce at most 10 errors; the average subject, however, only produced 1.32 errors (33 errors in all). Four errors were omissions, three were intrusions, and the rest were internal errors. An internal error is a response from the task's vocabulary which the subject had produced at the wrong time. Three of the pairs were supposed to contain equally available associates: A-B, B-C, and C-F. These pairs incurred 0, 7, and 3 forward errors, respectively, and 2, 6, and 2 backward errors. Hence, the total number of forward errors exactly equaled the total number of backward errors. The pairs D-A and E-B, however, which were supposed to be directional, incurred 3 and 1 forward errors, respectively, but 5 and 4 backward errors.

Response latencies are presented pair by pair. D-A was supposed to be directional, so a subject who did recall both associates was expected to recall A faster. Nineteen subjects recalled both associates correctly, and 14 subjects recalled A faster. The mean latency for D was 2.30 seconds, that for A, 1.83 seconds. The difference was statistically significant, t (18) = 2.14, $p < .05$.

Since B-C and C-F were supposed to be symmetrical, the forward and backward recalls were not to have differed in latency. The results confirmed this hypothesis. Eighteen subjects recalled both associates of B-C; 10 of them recalled the forward faster, and 8 recalled the backward faster. The mean latency of B's recall was 3.79 seconds and that of C's recall, 3.11 seconds. The difference was not statistically significant; t (17) = 1.11, $p > .10$. Twenty-one subjects recalled both associates of C-F, and only 9 of them recalled the forward faster. The mean latency of C's recall was 1.91 seconds and that of F's recall, 2.33 seconds. This difference was not statistically significant either; t (20) = −1.40, $p > .10$.

The two remaining pairs, A-B and E-B, introduced a special complication. When B was displayed for *backward* recall, A and E were both possible answers, and the subject tried to supply them both. The authors expected the subject to recall A and B with equal ease, but to take longer on E. (A-B was symmetrical, E-B was not.) This hypothesis was not confirmed. The competition in backward recall may have camouflaged a latency difference between A and E. Nineteen subjects supplied both backward recalls; 9 of them recalled A first and 10 recalled E first. The latencies of A or E, whichever occurred first, were recorded, and the mean A and E latencies of this listing were compared. The mean A latency was 3.57 seconds; the mean E latency was 2.92 seconds. This difference was not statistically significant; $t < 1.00$. The mean of the entire listing, 3.26 seconds, was significantly greater than the corresponding B latencies of the forward recalls; F (2, 72) = 3.44, $p < .05$. The B latencies of A-B's forward recall averaged 2.54 seconds, that of E-B's forward recall, 2.10 seconds.

Other Causes of Directionality

These results generally support the principle of associative symmetry, but the principle poses further problems. For example, most words of everyday English are so highly available that their AV values cannot differ materially; still, asymmetrics emerge in word-association norms: BEAUTIFUL strongly elicits WOMAN in the Minnesota norms (Russell & Jenkins, 1954), but WOMAN scarcely elicits BEAUTIFUL; AFRAID strongly elicits DARK, but DARK does not elicit AFRAID. The authors believe that the subject's "set" may create these asymmetries, and several sources of set are examined next.

First, the subject's task instructions may create directionalities. The subject may be instructed (or he may instruct himself) to supply, say, descriptive associates. Then NIGHT would strongly elicit DARK—not so much through verbal habits as through the subject's view of reality. There probably are no instructions, however, which can demand NIGHT so uniquely as a response to DARK.

Verbal behavior is always constrained in this way. If the task instructions do not specify constraints, then the rules of English grammar do. Thus, grammatical rules define a second set that may influence word associations. The subject has learned transitional probabilities from one grammatical class to another in spoken English, so he responds to an adjective with a noun and to a noun with a verb. Associations may still be symmetrical for practical matters of memory, but the set to be grammatical in free association may obscure these symmetries.

The authors have tried to simulate a grammatical set experimentally, but the attempt failed. The procedure of Experiment II was adopted, and the subject learned five pairs of associates. Each pair was embedded in meaningful sentences to give each associate grammatical form. The pair, VOLVAP-TAROP, for example, appeared as "A VOLVAP often TAROPS," "A VOLVAP is TAROPING," and so on. The sentence frames varied from trial to trial, but VOLVAP always stood as a noun and TAROP as a verb. Then free associations to every item were collected. Directional associations were expected in place of former symmetries. The response hierarchies, though, did not differ from those of Table 3 in any significant respect.

Third, an item's context may produce a set for directional associations. "Context" denotes other verbal conditions which enhance some associate's probability of occurring (Howes & Osgood, 1954). Say Items A and B elicit each other with some small probability. When A appears in Context C, though, it may elicit B more strongly. In other words, the compound stimulus, A + C, would elicit B more readily than A alone. On the other hand, contexts may not exist which could make B elicit A that strongly. So the compound stimulus would be a source of directionality.

The compound stimulus may keynote the way serial learning becomes directional. The earlier "chaining hypothesis" of serial learning asserted that each item of a serial list comes to elicit the next item. This hypothesis, however, has been amply refuted (e.g., Young, 1959). The principle of associative symmetry counters it still further: Since all serial items become available, the chaining hypothesis implies symmetrical associations between each pair of adjacent items. Item n would elicit items $n - 1$ and $n + 1$ with equal ease. Thus, serial learning ought to be much more arduous than it is. Although the serial task is still puzzling, the items that precede a given item seem to serve as a collective stimulus for it (Horowitz, Lippman, Norman, & McConkie, 1964). K alone does not elicit L in the alphabet, but a compound of preceding letters does. This hypothesis explains why one can recite the alphabet forwards so much more easily than backwards.

General Implications

The concepts of availability and associative symmetry help organize the data of human memory with appealing simplicity. The following implications illustrate this simplicity.

Meaningfulness and Familiarization

An item's availability must correlate with its meaningfulness (*m*) and with its exposure frequency (*n*). According to Noble (1952), a word's *m* grows as it acquires more associates; its *n* grows with each additional exposure. Of course, the operations that increase *m* and *n* also increase AV. Empirical research must still decide whether *m*, *n*, and AV are experimentally separable, but in this discussion they are assumed to be correlated at least.

Conclusions about availability can thus be applied to discussions of *m* and *n*, and vice versa. A PA task demands available *responses*, so it becomes easier when its responses are more meaningful or more familiarized. Several authors have already shown the critical importance of response meaningfulness (Cieutat, 1959; Cieutat, Stockwell, & Noble, 1958; Hunt, 1959; Kothurkar, 1963; Morikawa, 1959; and Weiss, 1958). These authors have examined different values of stimulus *m* combined with different values of response *m*. The learning rate was greatest when the responses were meaningful. In particular, some tasks made the subject associate a low-*m* item with a high-*m* item; the task was always easier when the high-*m* item served as response.

Response familiarity would seem to facilitate PA learning by making the responses more available. Experimental evidence, though, is equivocal. Part of the ambiguity is due to the welter of unstandardized procedures for increasing *n* (Bailey & Jeffrey, 1958; Cieutat, 1960; Gannon & Noble, 1961; Mandler & Campbell, 1957; Saltz, 1961; Schulz & Tucker, 1962; Weiss, 1958). Some experimenters have had the subject study the familiarized items briefly (e.g., Cieutat, 1960). Others have had him produce the items repeatedly by free recall (Horowitz & Larsen, 1963). Sometimes familiarization has introduced a special interference component into the subject's task (Horowitz & Larsen, 1963) and could camouflage the advantage of gained availability.

Backward Associations

Asch and Ebenholtz (1962a) have reviewed the data which show how prevalent backward associations are. Although forward recall usually surpassed backward recall, the responses in these studies had usually become more available than the stimuli. Given that backward recall occurs, though, it should occur most readily for available stimuli. Jantz and Underwood (1958) and Hunt (1959) have confirmed this claim. Hunt varied stimulus and response *m* in a PA task. First, the subject learned the PA task to mastery. Then each response was presented alone, and the subject had to supply its stimulus. Backward recall was best when stimulus *m* (hence stimulus AV) was high.

Paired-Associate Learning

Underwood et al. (1959) have analyzed paired-associate learning into two theoretical stages. One stage concerns the response's availability and the other, its association with the stimulus. The present results further testify to the power of their analysis.

Underwood et al. (1959) have not specified whether the response must be available to enter an association. The present authors believe not and assert that the stages are independent. Suppose two items, A and B, co-appear in some verbal context, and the subject does not produce them. The items would become associated, but both directions of the association would remain latent. The phenomenon would resemble latent learning.

The authors argue that item availabilities mainly affect performance. If A or B later grew in availability (after the association was formed), the A-B association would emerge in performance. This analysis would explain why performance is so much better on a recognition task than on a recall task. A recognition task provides the response alternatives and maximizes their availabilities.

Experiments with humans do not usually distinguish between learning and performance factors. But an association cannot emerge in the subject's behavior unless the response is available enough. Latent associations which exist in the nervous system might be called "unconscious" and might still affect the subject's behavior: They might contribute to the associative meaning of some concept, help shape some attitude, or have clinical importance, as described below.

Psychotherapy

Certain associations between unlabeled associates are especially prominent in psychotherapy. Let A and B denote two cognitions—A, one that develops when the subject perceives some subtle interpersonal relationship; and B, one that develops when the subject experiences an unpleasant feeling. The association A-B would then link one of these perceptions with one of these feelings. Perhaps, for example, a child perceives a subtle exchange of hostilities between his parents (A). Through identifying and through empathy, his perception comes to arouse an indescribable mixture of anxiety and anger (B). Such perceptions usually remain unlabeled (hence, unavailable), so A would be an unavailable stimulus. The child could not identify the source of his emotional discomfort since B, the feeling, does not elicit A, its source.

A psychotherapist might eventually hope to get a patient's feeling "worked through," extinguished, or modified by discrimination training. But as the first step, the stimulus must be identified. As long as it is not available, the behavior cannot be modified. Thus, a therapist has to heighten availabilities —a prerequisite to any therapeutic strategy he might later adopt. Also, according to this paper's view, the patient should be the one who labels the perceptions. Otherwise, the availabilities stay less than maximal.

Transfer

Positive transfer occurs when some task demands an association that the subject has already learned through contiguity or contextual similarity. Performance would remain imperfect, though, until all responses were available enough. Thus, some task might demand an association that had already been induced, but this association would remain latent until the response grew more available.

This interpretation can explain differences among some transfer paradigms. Horton and Hartman (1963) had the subject learn three PA tasks composed of low-frequency five-letter words. The first task was denoted A-B, the second task, B-C, and the third task was a mixture of A-C and C-A. The A-C pairs of the third task showed the greater transfer. Horton and Hartman inferred that "backward associations" are less effective mediators, but an alternate view exists: The first two tasks induced an association between A and C. Since Item C had been a response, it was the more available. And so the induced association was directional—from A, the less-available item, to C. The writers believe that equal transfer would have ensued if A and C had been equally available.

McCormack (1961) studied transfer between PA tasks of high-frequency words. First the subject learned a task denoted A-B, then one denoted C-B, and finally one denoted A-C. The resulting transfer was positive, so the authors viewed backward associations as effective mediators. But the stimuli of this experiment, high-frequency words, were available to start with; thus, the induced association A-C was overt and symmetrical. The writers believe that the transfer to C-A would have been equally positive.

Response competition causes negative transfer, but the two competing associates may emerge simultaneously. Murdock (1958) had three experimental groups learn an A-B task. Each group then transferred to a second task—one group to A-C, another to B-C, a third to C-A. All groups displayed negative transfer on the second task, and the interference was worst to C-A. (In Murdock's data the difference did not reach statistical significance, but in Goulet and Barclay's, 1965, replication it did.) Some explanation is still needed for the gross interference from A-B to C-A—since C, it seems, had just one associate. When the subject learned C-A, the A-B stimulus would have become more available, and the A-B association would have become symmetrical. Thus, B and C were both eliciting A, hence inducing an association between them. This induced association, C-B, would have then competed with C-A to produce the observed interference.

These principles can also help interpret recent results of Keppel and Underwood (1962) on backward associations. (Their results were replicated by Kausler & Kanoti, 1963.) First the subject learned an A-B list to one perfect trial. A denotes a nonsense syllable and B, an adjective. Then the subject spent 15 trials learning a second task; four second tasks were examined. Finally, the subject was tested for backward recall of the first task. He was shown each response, and he had to supply its stimulus.

TABLE 4

DATA OF KEPPEL AND UNDERWOOD (1962) ORDERED AND INTERPRETED

First task	Second task	Mean B-A recall	Is B free of other associates?	Does A stay available?
A-B	None	5.80	Yes	Unknown
A-B	A-C	5.05	Yes	Yes
A-B	C-D	4.60	Yes	No
A-B	A-Br	3.40	No	Yes
A-B	C-B	1.80	No	No

The results of this experiment are presented in Table 4 according to the subject's second task. Table 4 manages to rank order these results by establishing two facts in Columns 4 and 5. Column 4 tells whether B was free of other associates. The prognosis was poor when B did appear in the second task. Column 5 tells whether A kept being available during the second task. Item A clearly did stay available if it appeared in the second task. Thus, the net prognosis was best for backward recall when that recall followed A-C; it was worst after C-B. If greater weight is attached to Column 4, Keppel and Underwood's (1962) results make sense.

"Availability" may help relate research in verbal learning to general psychology. Brown and Lenneberg (1954) have shown that a color whose name is more "codable" is better remembered. Their study was designed to test Whorf's hypothesis, but its interpretation was partly confounded. The most codable color names (high-frequency words) are probably used to label the most frequent sensory experiences. Thus, word frequencies and sensation frequencies probably covary. And Brown and Lenneberg's (1954) result may reflect a perceptual phenomenon instead of a linguistic one. Codability seems to be related to availability, though, so a color's *name* might be made more codable by the methods of this paper. Thus, these methods might help to clarify and evaluate Whorf's hypothesis.

REFERENCES

ASCH, S. E., & EBENHOLTZ, S. M. The principle of associative symmetry. *Proceedings of the American Philosophical Society,* 1962, **106**, 135–163. (a)

ASCH, S. E., & EBENHOLTZ, S. M. The process of free recall: Evidence for nonassociative factors in acquisition and retention. *Journal of Psychology,* 1962, **54**, 3–31. (b)

ASCH, S. E., & LINDNER, M. A note on "strength of association." *Journal of Psychology,* 1963, **55**, 199–209.

BAILEY, J. H., & JEFFREY, W. E. Response strength and association value in stimulus predifferentiation. *Psychological Reports,* 1958, **4**, 715–721.

BROWN, R., & LENNEBERG, E. H. A study in language and cognition. *Journal of Abnormal and Social Psychology,* 1954, **49**, 454–462.

CIEUTAT, V. J. Supplementary report: Stimulus and response meaningfulness (m') in paired-associate learning by hospitalized mental patients. *Journal of Experimental Psychology,* 1959, **58**, 490.

CIEUTAT, V. J. Differential familiarity with stimulus and response in paired-associate learning. *Perceptual and Motor Skills,* 1960, **11**, 269–275.

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m') in paired-associate learning. *Journal of Experimental Psychology,* 1958, **56**, 193–202.

GANNON, D. R., & NOBLE, G. E. Familiarization (n) as a stimulus factor in paired-associate verbal learning. *Journal of Experimental Psychology,* 1961, **62**, 14–23.

GOULET, L. R., & BARCLAY, A. Supplementary report: A comparison of paired-associate transfer effects between the A-B, C-A and A-B, B-C paradigms. *Journal of Experimental Psychology,* 1965, **70**, 537–538.

HOCKETT, C. F. *A course in modern linguistics.* New York: Macmillan, 1958.

HOROWITZ, L. M., BROWN, Z. M., & WEISSBLUTH, S. Availability and the direction of associations. *Journal of Experimental Psychology,* 1964, **68**, 541–549.

HOROWITZ, L. M., & LARSEN, S. R. Response interference in paired-associate learning. *Journal of Experimental Psychology,* 1963, **65**, 225–232.

HOROWITZ, L. M., LIPPMAN, L. G., NORMAN, S. A., & MCCONKIE, G. W. Compound stimuli in paired-associate learning. *Journal of Experimental Psychology,* 1964, **67**, 132–141.

HORTON, D. L., & HARTMAN, R. R. Verbal mediation as a function of associative directionality and exposure frequency. *Journal of Verbal Learning and Verbal Behavior,* 1963, **1**, 361–364.

HOWES, D., & OSGOOD, C. On the combination of probabilities in linguistic contexts. *American Journal of Psychology,* 1954, **67**, 241.

HUNT, R. G. Meaningfulness and articulation of stimuli and responses in paired-associate learning and stimulus recall. *Journal of Experimental Psychology,* 1959, **57**, 262–267.

HUNTER, W. S. The delayed reaction in animals and children. *Behavior Monographs,* 1913, **2**(1).

JANTZ, E. M., & UNDERWOOD, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *Journal of Experimental Psychology,* 1958, **56**, 174–179.

KAUSLER, D. A., & KANOTI, G. A. R-S learning and negative transfer effects with a mixed list. *Journal of Experimental Psychology,* 1963, **65**, 201–205.

KEPPEL, G., & UNDERWOOD, B. J. Retroactive inhibition of R-S associations. *Journal of Experimental Psychology,* 1962, **64**, 400–404.

KJELDERGAARD, P. M., & HORTON, D. L. An experimental analysis of associative factors in stimulus equivalence, response equivalence and chaining paradigms. *Studies in verbal behavior.* Report No. 3, 1960, University of Minnesota, Grant No. G-18690, National Science Foundation.

KOTHURKAR, V. K. Effect of stimulus-response meaningfulness on paired-associate learning and retention. *Journal of Experimental Psychology,* 1963, **65**, 305–308.

MANDLER, G., & CAMPBELL, E. H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. *Journal of Experimental Psychology*, 1957, **54**, 269–273.

McCORMACK, P. D. Backward mediated positive transfer in a paired-associate task. *Journal of Experimental Psychology*, 1961, **61**, 138–141.

McNEILL, D. The origin of associations within the same grammatical class. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 250–262.

MORIKAWA, Y. Functions of stimulus and response in paired-associate verbal learning. *Psychologia*, 1959, **2**, 41–56.

MURDOCK, B. B. Backward associations in transfer and learning. *Journal of Experimental Psychology*, 1958, **55**, 111–114.

NOBLE, C. E. An analysis of meaning. *Psychological Review*, 1952, **59**, 421–430.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Technical Report No. 11, 1954, University of Minnesota, Contract N8-onr-66216, Office of Naval Research.

SALTZ, E. Response pretraining: Differentiation or availability? *Journal of Experimental Psychology*, 1961, **62**, 583–587.

SCHULZ, R., & TUCKER, I. F. Supplementary report: Stimulus familiarization in paired-associate learning. *Journal of Experimental Psychology*, 1962, **64**, 549–550.

UNDERWOOD, B. J. Articulation in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 146–149.

UNDERWOOD, B. J., RUNQUIST, W. N., & SCHULZ, R. W. Response learning in paired-associate lists as a function of intralist similarity. *Journal of Experimental Psychology*, 1959, **58**, 70–78.

WEISS, R. L. The role of association value and experimentally produced familiarity in paired-associate verbal learning. (Unpublished doctoral dissertation, University of Buffalo, 1958) *American Psychologist*, 1958, **13**, 386. (Abstract)

YOUNG, ROBERT K. A comparison of two methods of learning serial associations. *American Journal of Psychology*, 1959, **72**, 554–559.

(Received August 31, 1964)

Psychological Review
1966, Vol. 73, No. 1, 16–43

HABITUATION:

A MODEL PHENOMENON FOR THE STUDY OF NEURONAL SUBSTRATES OF BEHAVIOR[1]

RICHARD F. THOMPSON AND WILLIAM ALDEN SPENCER[2]

University of Oregon Medical School

The recent habituation literature is reviewed with emphasis on neurophysiological studies. The hindlimb flexion reflex of the acute spinal cat is used as a model system for analysis of the neuronal mechanisms involved in habituation and sensitization (i.e., dishabituation). Habituation of this response is demonstrated to follow the same 9 parametric relations for stimulus and training variables characteristic of behavioral response habituation in the intact organism. Habituation and sensitization appear to be central neural processes and probably do not involve presynaptic or postsynaptic inhibition. It is suggested that they may result from the interaction of neural processes resembling "polysynaptic low-frequency depression," and "facilitatory afterdischarge." "Membrane desensitization" may play a role in long-lasting habituation.

Perhaps the most fundamental problem in physiological psychology is the elucidation of neuronal mechanisms underlying alterations in behavior. This implies prediction of behavioral changes from descriptions in terms of patterns of synaptic interactions among identifiable populations of neurons. Two general strategies in vogue today are (*a*) correlational studies relating changes in various neuronal responses of the intact organism to changes in behavior (cf. Morrell, 1961), and (*b*) analytic studies of relatively long-term changes induced in simplified neuronal systems such as the monosynaptic reflex (cf. Eccles, 1964). At present, phenomena measured in the intact organism cannot be analyzed in terms of neuronal mechanisms. Analytic studies, while utilizing simpler systems, generally deal with changes whose relationships to alterations in the behavior of the intact organism are not always clear.

The solution to this impasse would seem to be a type of relatively long-lasting and nontrivial alteration in behavior as a result of training which has identical characteristics both in the intact organism and in simplified preparations, where some degree of neuronal analysis is possible. There has been a resurgence of interest in recent years in the phenomena of habituation, due in part to studies of neural response habituation (e.g., Hernández-Peón, Jouvet, & Scherrer, 1957; Sharpless & Jasper, 1956). Behavioral response habituation appears to satisfy both the requirements of behavioral significance and occurrence in simplified preparations. A wide variety of responses of the intact organism habituate (Harris, 1943), as

---

[1] Supported in part by NIH Research Grants NB-03494 and B-2161 and by NIH Research Career Awards 1-K-NB-19485 and K3-MH-6650. The authors are indebted to John M. Brookhart, Judson S. Brown, Seth Sharpless, and Eric Kandel for helpful comments and suggestions.

[2] W. A. Spencer's current address: Department of Physiology, New York University School of Medicine.

do reflex responses of the chronic spinal animal (Prosser & Hunter, 1936). Sufficient information has been developed about the neural organization of the spinal cord to permit some degree of analysis of reflex mechanisms (e.g., Eccles, 1964). This paper will discuss recent studies of habituation with emphasis on neurophysiological findings. We will attempt to demonstrate that habituation of the spinal flexion reflex is identical in characteristics to habituation of more complex responses of the intact organism and is at least partially amenable to analysis in terms of neuronal mechanisms. Original data to be reviewed were all obtained using reflex activity of a "model" mammalian system: the acute spinal cat.

DEFINITIONS

Harris (1943), in his careful and comprehensive review of the earlier literature, defined habituation operationally to be "response decrement as a result of repeated stimulation [p. 385]." Humphrey's (1933) earlier definition, although expressed in less-behavioral terms, is substantially the same. Response decrement resulting from very rapid stimulation (e.g., 500/second) which might be due to invasion of the neuronal relative refractory period has different characteristics and is best excluded, as are decrements resulting from trauma, growth, aging, etc. Habituation can usually be distinguished from the latter phenomena because it is reversible: Habituated responses exhibit spontaneous recovery. The only authority to diverge from Harris' characterization of habituation is Thorpe (1956), who regards it as a "relatively *permanent* waning of a response as a result of repeated stimulation [p. 54; italics ours]." It will be seen that the time course for spontaneous recovery of an habituated response depends on a great many variables; hence distinctions in terms of recovery time would seem somewhat arbitrary. Certainly Harris' definition has gained widespread acceptance. Several authors have suggested a distinction between "short-term" and "long-term" habituation processes (Shapless & Jasper, 1956; Sokolov, 1955).

Two peripheral mechanisms have often been assumed either explicitly or implicitly to account for response decrement: alterations in receptors or in effectors. These may be given more specific operational definitions which permit empirical evaluation. In this paper, a decrease in receptor activity will be called *receptor adaptation*, and a decrease limited to the effector response (including both effector and neuroeffector junction) will be termed *effector fatigue*. Following Harris (1943), response decrements that can be accounted for entirely by these mechanisms will not be considered habituation.

If a common-denominator definition of learning such as "change in behavior under conditions of practice" is adopted, habituation must be included as an aspect of learning. Habituation can be distinguished from many phenomena of learning in that the change in behavior is a decrease in response strength. It is customary, however, to assume that the response exhibiting habituation is itself unlearned (Harris, 1943). James (1890) and Konorski (1948) have proposed a more inclusive term, *plasticity*, to include habituation, conditioning, and other aspects of behavioral change. Our general definition of habituation may also apply to a variety of other terms: adaptation, accommodation, fatigue, inhibition, negative learning, extinction, stimulus satiation, etc. Habituation will be used here to stand for whichever term the reader prefers. Martin (1964) has recently proposed that such diverse constructs as response habituation, evolutionary adaptation, and photochemical dark adaptation have a common conceptual basis. The utility of this approach is unclear.

PARAMETRIC CHARACTERISTICS OF HABITUATION

The extensive research on behavioral response habituation in the intact vertebrate has demonstrated a number of parametric relations for stimulus and training variables. Although not previously enumerated as such, examples of all of these may be found in the earlier literature for various reflex responses. Some of the more recent observations, particularly those dealing with types of responses not covered in Harris' (1943) review, will be noted in the brief summary that follows.

1. *Given that a particular stimulus elicits a response, repeated applications of the stimulus result in decreased response (habituation). The decrease is usually a negative exponential function of the number of stimulus presentations.*

Examples of response habituation can probably be found in essentially *all* behavioral studies where a stimulus is regularly presented. In earlier experiments devoted to habituation, per se, parametric characteristics were studied for a variety of responses (cf. Harris, 1943) ranging from postrotatory nystagmus (Griffith, 1920; Wendt, 1951) to startle (Prosser & Hunter, 1936) and galvanic skin response (GSR—Davis, 1934). With the possible exception of the "knee jerk" reflex (Lombard, 1887; Prosser & Hunter, 1936) habituation was a consistent finding, usually exhibiting an exponential course. The recent habituation literature contains many investigations of various responses which can be subsumed under the general heading of response to novel stimulation, ranging in complexity from simple startle through variously defined orienting reflexes to curiosity, stimulus satiation, and play behavior (Berlyne, 1950, 1960; Bindra, 1959; Butler & Harlow, 1954; Davis, Buchwald, & Frankmann, 1955; Dember, 1961; Dykman, Reese, Galbrecht, & Thomasson, 1959; Engen & Lipsitt, 1965; Glanzer, 1953; Melzak, 1961; Montgomery, 1953; Robinson & Gantt, 1947; Rodgers, Melzak, & Segal, 1963; Scholander, 1960; Sokolov, 1960; Welker, 1956, 1961). Interestingly, if a single novel stimulus is repeated regularly, habituation of approach behavior follows an exponential course (Glanzer, 1953). Unlearned behavior sequences of birds elicited by certain categories of stimuli have also been found to habituate under some conditions (Hinde, 1954; Thorpe, 1956). Finally, much of the extensive research on response decrement or fatigue under conditions of work (Bartley & Chute, 1947; Robinson, 1934; Solomon, 1948) can be considered within the framework of habituation.

2. *If the stimulus is withheld, the response tends to recover over time (spontaneous recovery).*

Spontaneous recovery is reported in most of the studies noted above and has come to be the most common method of demonstrating that a given response decrement is an example of habituation (Harris, 1943). The time course of spontaneous recovery is markedly influenced by many variables and is not necessarily characteristic of a given response. Thus the habituated startle response to sound in the intact rat may recover in 10 minutes (Prosser & Hunter, 1936) or fail to recover in 24 hours,[3] depending upon details of testing. Consequently any categorization of types of habituation based solely on recovery time is likely to be somewhat artificial.

3. *If repeated series of habituation training and spontaneous recovery are given, habituation becomes successively more rapid (this might be called potentiation of habituation).*

Humphrey (1933) noted this effect in his studies on turtle leg withdrawal to shell tap. Konorski (1948) describes it for the orienting response, and it has been described in many studies where repeated habituation series were given (e.g., Davis, 1934).

4. *Other things being equal, the more rapid the frequency of stimulation, the more rapid and/or more pronounced is habituation.*

---

[3] J. S. Brown, personal communication, 1964.

Numerous examples of this were noted in the earlier reflex studies (Harris, 1943) as well as in more recent work on stimulus satiation and curiosity (Glanzer, 1953; Welker, 1961). The effect occurs in terms of real time course and occurs within certain limits in terms of number of trials as well.

5. *The weaker the stimulus, the more rapid and/or more pronounced is habituation. Strong stimuli may yield no significant habituation.*

This relationship is characteristic of most types of responses ranging from simple reflexes (Harris, 1943) to complex exploratory behavior (Welker, 1961).[4]

6. *The effects of habituation training may proceed beyond the zero or asymptotic response level.*

Additional habituation training given after the response has disappeared or reached a stable habituated level will result in slower recovery. Although relatively few experiments have studied "below-zero" habituation as such (Humphrey, 1933; Prosser & Hunter, 1936; Wendt, 1931), the observations may be viewed as an extension of the relationship between number of stimulus presentations and degree of habituation. Zero response level is of course to some degree dependent upon the particular response measures used.

7. *Habituation of response to a given stimulus exhibits stimulus generalization to other stimuli.*

Coombs (1938) demonstrated generalization of GSR habituation to different types of auditory stimulation, and Porter (1938) demonstrated cross-modal generalization of the habituated GSR for light and tone stimuli. Mowrer (1934) showed some generalization of postrotatory nystagmus habituation in the pigeon. In a recent study, Crampton and Schwam (1961) reported generalization of optic nystagmus habituation in the cat to different degrees of angular acceleration.

8. *Presentation of another (usually strong) stimulus results in recovery of the habituated response (dishabituation).*

This phenomenon appears to be as ubiquitous as habituation itself and is commonly used to demonstrate that habituation has occurred. Pavlov (1927) was perhaps the first to describe this process (i.e., disinhibition) in relation to an extinguished conditioned response (CR), but also applied it to the habituated orienting response. Humphrey (1933) studied dishabituation extensively in lower vertebrates. Essentially all responses of mammals that can be habituated can also be dishabituated (Harris, 1943). It is not always necessary for the dishabituatory stimulus to be strong. In fact Sokolov (1960) and Voronin and Sokolov (1960) reported that a *decrease* in the intensity of an auditory stimulus results in dishabituation of the habituated orienting response in humans. Dishabituation, viewed as neutralization of the process of habituation (Humphrey, 1933), has been perhaps the most important method of distinguishing between habituation and "fatigue."

9. *Upon repeated application of the dishabituatory stimulus, the amount of dishabituation produced habituates (this might be called habituation of dishabituation).*

Most studies of dishabituation (see above) have noted its habituation. Lehner (1941) has done the most careful parametric studies, showing that habituation of dishabituation follows a negative exponential course for the startle response in the rat and the abdominal reflex in man. More recently, Hagbarth and Kugelberg (1958) and Hagbarth and Finer (1963) verified and extended Lehner's findings for the abdominal and leg flexion reflexes in humans. Crampton and Schwam (1961) have shown that dishabituation of postrotatory nystagmus in the cat by auditory or cutaneous stimuli habituates in a similar fashion.

In reviewing the behavioral habituation literature, it is striking to find virtually complete agreement on the parametric characteristics of the phenomenon in such a wide variety of animals and responses. *These nine*

---

[4] Postrotary optic nystagmus may be an exception in that under some conditions the degree of habituation is directly related to velocity of rotation (G. Crampton, personal communication, 1964).

common characteristics may consequently serve as the detailed operational definition of habituation, replacing the more general definition given above. The extent to which any other response decrements satisfy these characteristics will thus determine whether they can be called habituation.

NEUROPHYSIOLOGICAL STUDIES OF HABITUATION

Electroencephalogram (EEG) Arousal

The EEG arousal response—a rapid transition from slower waves to fast low-voltage activity, usually following sudden stimulation—was observed to decrease upon repeated stimulation in early studies both in animals (e.g., Rhineberger & Jasper, 1937) and man (Knott & Henry, 1941). The first investigation of this effect, per se, was the now classic study by Sharpless and Jasper (1956). Using repeated presentations of brief tones they found that cortical EEG arousal of the sleeping cat becomes shorter and finally disappears. After cessation of stimulation, the arousal response exhibited spontaneous recovery over a period of minutes to hours. Although not studied parametrically, examples of all nine relations listed above for behavioral responses are to be found here and in subsequent confirmatory studies (Caspers, Lerche, & Greuter, 1958; Roig & Sommer-Smith, 1959; Segundo, Roig, & Sommer-Smith, 1959). Apelbaum, Silva, Frick, and Segundo (1960) noted a more restricted generalization gradient when using waking cats than Sharpless and Jasper found with sleeping animals. The human "alpha blocking" response has been shown to habituate to tactile, auditory, and visual stimulation (Sokolov, 1960; Voronin & Sokolov, 1960). Cortical EEG arousal also exhibits habituation, spontaneous recovery, and dishabituation when elicited by electrical stimulation of the mesencephalic reticular formation (Glickman, 1958; Glickman & Feldman, 1961; Roig & Sommer-Smith, 1959). Habituation of electrographic arousal has also been reported for thalamus and reticular formation (Sharpless & Jasper, 1956), hippocampus (Green & Arduini, 1954), and olfactory bulb (Adrian, 1950; Hernández-Peón, Lavin, Alcocer-Cuarón, & Marcelin, 1960).

Sensory Evoked Potentials

A number of investigators have reported habituation of click- or tone-evoked responses in various regions of the central nervous system (CNS) ranging from the cochlear nucleus to the auditory cortex and mesencephalic reticular formation in the waking animal (Galambos, Sheatz, & Vernier, 1956; Gershuni, Kozhevnikov, Maruseva, Avakyan, Radionova, Altman, & Soroko, 1960; Hernández-Peón et al., 1957; Hernández-Peón & Scherrer, 1955; Jouvet & Hernández-Peón, 1957). Worden and Marsh (1963) have emphasized the importance of controlling stimulus intensity, demonstrating statistically reliable changes in response, and conducting other necessary control procedures which were not always observed in the studies listed above. Experiments satisfying all these methodological criteria have not, for the most part, found any consistent or significant evidence of habituation of auditory evoked responses in the waking animal (Huttenlocher, 1960; Marsh, McCarthy, Sheatz, & Galambos, 1961; Sharpless & Jasper, 1956; Worden & Marsh, 1963). However Marsh and Worden (1964) have recently reported decreases in click-evoked response amplitudes from the auditory cortex when the EEG was desynchronized, even in the absence of any consistent response changes at the cochlear nucleus. In a recent and careful study, Webster, Dunlop, Simons, and Aitkin (1965) reported a rapid initial decrease in the amplitude of sound-burst-evoked responses of the cochlear nucleus that occurred during the first 30 seconds of stimulation, if the sound (20-millisecond duration) was presented 1/second or *faster*. No decreases occurred with slower rates of stimulation. (The experiments described above generally presented stimuli slower than 1/second and examined responses over periods of many minutes or hours.) Auditory evoked responses of the reticular formation do appear to habituate in the sleeping animal (Huttenlocher, 1960). Auditory evoked cortical association responses do not habituate (Shaw & Thompson, 1964); in fact the amplitude of this response increases during the first few presentations of a click (Thompson & Shaw, 1965). Head position in the sound field is a powerful and consistent determiner of evoked-response amplitude (Marsh, Worden, & Hicks, 1962). Contractions of middle-ear muscles have been implicated in some studies of auditory evoked response habituation (Guzmán-Flores, Alcaraz, & Harmony, 1960; Hugelin, Dumont, & Paillas, 1960), but denied by other workers (Altman, 1960; Moushegan, Rupert, Marsh, & Galambos, 1961). Further discussion of methodological problems is given in the careful studies by Marsh et al. (1962) and Worden and Marsh (1963). (See also the discussion by Deutsch, 1962.)

Many reports have noted habituation of evoked responses from various regions of the visual system to repeated flash stimulation (Cavaggioni, Giannelli, & Santibañez, 1959; Hernández-Peón, Guzmán-Flores, Alcaraz, & Fernandez-Guardiola, 1958; Mancia, Meulders, & Santibañez, 1959; Palestini, Davidovich, & Hernández-Peón, 1959). Pupil diameter apparently was not controlled in many of these early studies. When the eye is atropinized and an artificial pupil used, the evidence for visual evoked response habituation is inconsistent. Fernandez-Guardiola, Roldán, Fanjul, and Castells (1961) found no habituation of responses to flash under these conditions recording from the optic tract and lateral geniculate body and inconsistent changes from the visual cortex: Some preparations showed partial decreases, particularly of the later components of the evoked response, and others showed no decreases. Naquet, Regis, Fischer-Williams, and Fernandez-Guardiola (1960) noted that flash-evoked responses of the visual cortex decreased when pupil diameter increased (concomitant with EEG desynchronization), but did not change when an artificial pupil was used. Affanni, Mancia, and Marchiafava (1962) reported no reduction in responses of the visual cortex to a flash repeatedly presented for hours when an artificial pupil was used. On the other hand, Garcia-Austt (1963) reported decreases in scalp averaged evoked responses to flash in humans with artificial pupils, and decreases of averaged visual evoked responses from the rat after removal of the iris. Responses evoked by somatic sensory stimulation have been reported to habituate (Hernández-Peón, 1960; Hernández-Peón, Scherrer, & Velasco, 1956), and to exhibit no habituation (Santibañez, Trouche, & Albe-Fessard, 1963).

Conclusion

The evidence for EEG arousal habituation is clear and consistent; it exhibits essentially all parametric relations characteristic of behavioral response habituation. Studies of evoked responses, on the other hand, permit only one safe generalization at present:

There is as yet no consistent agreement regarding the occurrence of habituation in sensory systems of the waking brain. Where habituation has been reported, it may be the result of decreases in effective stimulus intensity due to alterations in head position, middle-ear muscle contractions, pupilary constriction, etc., or an indirect result of changes in state of EEG arousal. The issue can only be settled by further research.

EFFECTS OF NEURAL LESIONS ON BEHAVIORAL RESPONSE HABITUATION

Although the majority of studies described here were not planned as parametric investigations of the effects of lesions on behavioral response habituation, per se, many were so designed that such effects could be noted. The orienting response exhibits markedly reduced habituation following chronic total decortication in dogs (Konorski, 1948; Lebidinskaia & Rosenthal, 1935) and chronic decerebration in cats (Bard & Macht, 1958). Lesions of auditory cortex in the cat reduce the extent of intersession habituation of orientation to sound (Thompson & Welker, 1963). Activity level of monkeys with Cortical Area 9 lesions exhibits less habituation than that of normal animals (French, 1959; French & Harlow, 1955). Frontal or temporal lesions in monkeys reduce the amount of visual exploration and decrease the extent of habituation in repeated testing sessions (Butler & Harlow, 1954; Symmes, 1959, 1963). Frontal lesions have a comparable effect in the cat (Hagamen, Lance, & Ungewitter, 1959). Damage to frontal cortex decreases the extent of habituation of the heart-rate response to the cold-pressor test in man and the rat (Glaser & Griffin, 1962; Griffin, 1963). Schwartzbaum, Wilson, and Morrissette (1961) report less-rapid habituation to novel stimuli following bilateral amygdalectomy. Pribram (1963) reports unpublished results by Kimble and Bagshaw showing comparable effects of amygdalectomy on GSR response to novel stimulation and emphasizes the importance of this structure in habituation.

Response alternation or variability in maze behavior can be related to habituation of response to novel stimuli (see below; also Glanzer, 1953; Welker, 1961). Partial cortical lesions in the rat reduce the tendency to vary movement through a simple maze (Krechevsky, 1937a, 1937b, 1937c). Frontal lesions are more effective than parietal or occipital lesions (Morgan & Wood, 1943). Roberts, Dember, and Brodwick (1962) found that bilateral ablation of the hippocampus resulted in less habituation of T-maze responses and less alternation behavior.

Results of these studies, involving various responses and many different neural structures, are in essentially complete agreement. *Brain damage tends to produce a reduction in the extent of behavioral response habituation.* There is some degree of specificity in that damage to structures related to particular responses tends to have a greater effect on habituation. It would seem that habituation may be mediated, at least in part, by whatever neural structures are most involved in a particular response, suggesting that common neuronal mechanisms may be involved.

SPINAL REFLEXES

Sherrington noted a decrement of the digital flexion of the monkey in 1898, but his experiments on the acute spinal dog (1906) were the first to analyze habituation of spinal reflexes in detail. Although he used the term fatigue, he was careful not to imply any necessary meaning other than response decrement by the term. He found that continuous or repeated mechanical or electrical stimulation of the appropriate skin region resulted in a decrement of the scratch reflex and the flexion reflex, the former decreasing more rapidly. Both exhibited spontaneous recovery. Several lines of evidence suggested that both receptor adaptation and muscle fatigue could be ruled out. Indirect evidence from a variety of reflex-interaction studies led Sherrington to conclude that the "final common path," that is, the motor neurons, were not fatigued. Prosser and Hunter (1936) completed a careful and extensive study of habituation and dishabituation of the flexion reflex in the chronic spinal rat, using electrical or mechanical stimulation of the legs, tail, and hindlimb nerves. They demonstrated habituation, spontaneous recovery, and dishabituation by strong extraneous stimulation. Weak stimuli led to more rapid habituation than strong stimuli. In one experiment they showed that no habituation occurred in the knee-jerk reflex of a spinal cat. Lehner (1941) demonstrated that the duration of dishabituation of the flexion reflex in the spinal rat habituated.

In conditioning studies of the acute spinal dog, Shurrager and Culler (1941) and Shurrager and Shurrager (1941) demonstrated habituation of the leg flexion reflex to tail stimulation. More recently, Nesmeianova (1957) has demonstrated habituation, spontaneous recovery, and dishabituation of the scratch reflex and the flexion reflex in chronic spinal dogs. Habituation below zero was marked; the response sometimes required days or weeks to recover after long periods of training. Kozak, McFarlane, and Westerman (1962) have reported similarly long-lasting habituation of flexion and scratch reflexes in the chronic spinal kitten. Habituation of the scratch reflex has also been reported for the chronic spinal frog (Afelt, 1963). Lethlean (1965) has presented evidence suggesting that the flexion reflex may continue to habituate for many minutes after the stimulus has been withdrawn in the chronic spinal rat. Very recent experiments by Buchwald, Halas, and Schramm (1964, 1965) reported habituation of single-unit discharges of motor neurons recorded from ventral root fibers in the acute spinal cat, and a preliminary summary of our own work has already appeared (Spencer, Thompson, & Neilson, 1964).

REFLEX HABITUATION IN THE ACUTE SPINAL CAT

Studies reviewed above indicated that habituation of the flexion reflex in the acute mammalian spinal preparation appears to exhibit some of the parametric relations characteristic of behavioral response habituation in the intact organism. Since a good deal is known about the synaptic organization of the spinal cord, this preparation was chosen as a model system in which to study detailed characteristics of habituation and the possible neuronal mechanisms involved.

The preparation used in these experiments was the acute unanesthetized decerebrate cat with spinal transection at T-12. Procedures are described in detail in papers (in preparation) devoted to analyses of neuronal mechanisms. Data given here are all in terms of isometric "twitch" contraction strength of the tibialis anterior muscle, a hindlimb flexor participating in the generalized limb flexion reflex to afferent stimulation. Other muscles were also studied to insure generality of observations. The flexion reflex is polysynaptic, at least one interneuron intervening between the incoming afferent fibers and the output motor neurons. Dishabituating stimuli were either electrical or natural, that is, pinch. The basic phenomena of habituation and spontaneous recovery were observed on all of the more than 100 cats studied. Spontaneous recovery was used as the criterion that habituation had occurred rather than response decrement due to deterioration or various artifacts. Although not shown in many of the figures, spontaneous recovery was always obtained following every habituation series.

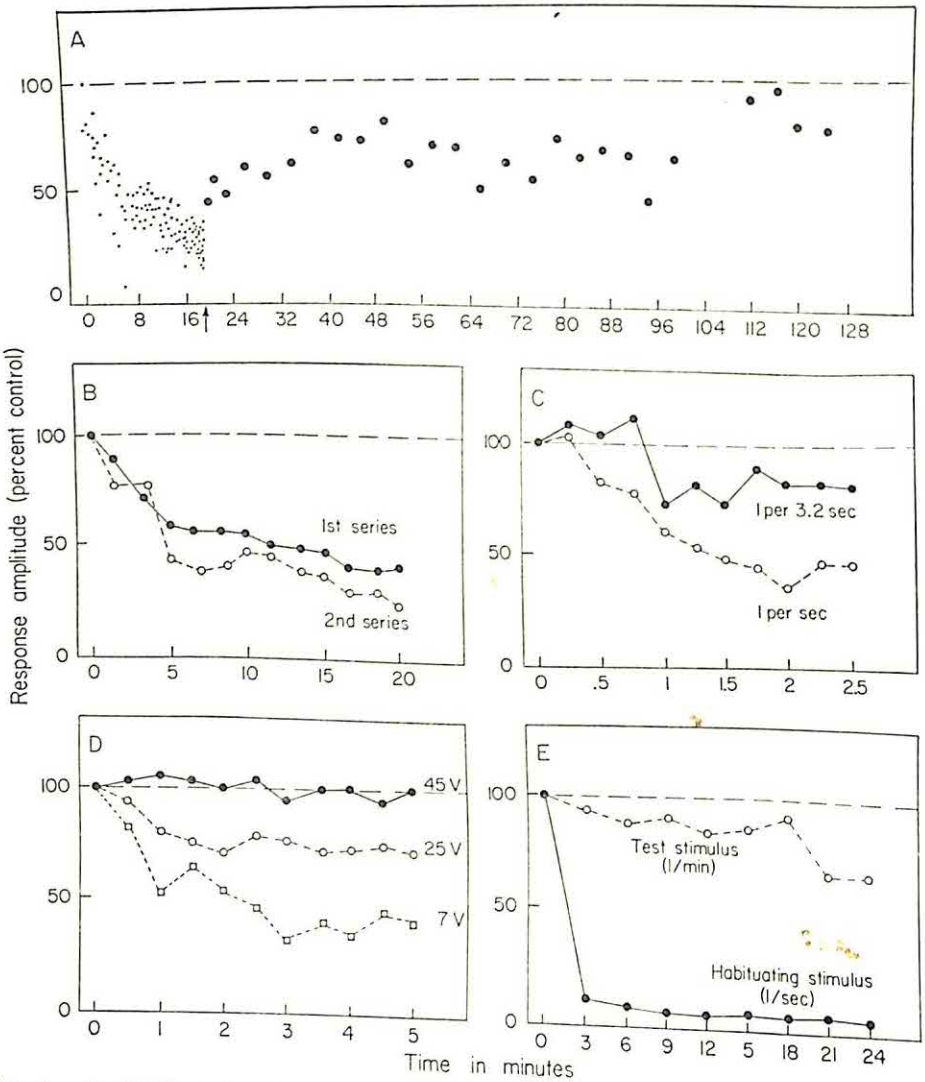

FIG. 1. A: Habituation (zero minutes to arrow) and spontaneous recovery (arrow to 128 minutes) of the hindlimb flexion reflex of the spinal cat in response to repeated skin shocks. (Stimuli were brief trains of shocks—5 in 50 milliseconds—delivered every 10 seconds during habituation and every 3 minutes during spontaneous recovery, except for a 12-minute period of no stimuli at about 100 minutes. In this and all subsequent figures the response measured is tension developed by contraction of the tibialis anterior muscle, expressed as a percentage of mean initial-control response amplitude. B: Effect of repeated habituation and spontaneous-recovery series on degree of habituation. Response recovered to control level following first habituation series and was then rehabituated—second series. Conditions as in A. Data are averages of 10-trial blocks. C: Effect of stimulus frequency on habituation. Single shocks given 1/3.2 seconds in one habituation series and 1/second in the other to the saphenous nerve. Data are averages of 10-trial blocks. D: Effect of stimulus intensity on habituation. Brief trains of shocks, as in A, were delivered every 10 seconds to the saphenous nerve with spontaneous recovery allowed after each series. Voltages refer to output of stimulator and were attenuated, but in the same ratios, when delivered to the nerve. Data averaged over 3-trial blocks. E: Stimulus generalization of habituation. Single shocks to two separate branches of the saphenous nerve. The habituating stimulus to one branch was given 1/second, and the test stimulus to the other branch was given 1/minute. Data are averages over 3-trial blocks for response to the test stimulus and averages over the same periods of time for response to the habituating stimulus.)

In all experiments, initial-control response amplitudes were first determined by test stimuli delivered no more frequently than 1/minute. Subsequent response amplitudes are plotted in terms of percentage of mean control level. Habituating stimuli were either single shocks (.01–.03-millisecond duration) or brief trains of five such shocks delivered over a 50-millisecond period. Although trains generally gave somewhat more rapid and less variable habituation, the parametric relations of all other variables were comparable for both. Stimuli were delivered to skin or to afferent nerves.

Characteristic habituation and spontaneous recovery of the spinal flexion reflex are shown in Figure 1A. Brief trains of shocks were delivered 1/10 seconds to skin during habituation (0 minutes to arrow) and 1/3 minute during recovery. Individual responses are given to illustrate normal variability. Note the exponential form of habituation and the long and variable time course of spontaneous recovery. Although the time courses of habituation and spontaneous recovery are dependent on the many variables described below, 40 separate series involving a variety of conditions were selected at random from our experiments to determine average values. The mean habituation time was 9.4 minutes (range 1.2–33 minutes), and the mean recovery time was 25.7 minutes (range .5 minute–108 minutes).

The effect of repeated habituation-spontaneous recovery series on habituation is illustrated in Figure 1B. Habituation and recovery conditions are the same as in 1A, and the data shown are averages for 10-trial blocks. The second series exhibits significantly more habituation than the first (p <.01; sign test).

Figure 1C shows the effect of stimulus frequency. Here single shocks were delivered either at 1/second or 1/3.2 seconds to the saphenous nerve. Data are averages over 16-second blocks of time. Habituation is significantly greater with the more rapid stimulus (p <.01; sign test).

The effect of stimulus intensity on habituation is indicated in Figure 1D. Stimuli were brief trains of shocks to the saphenous nerve every 10 seconds at the indicated voltages. Data are averages of 3-trial blocks. There is no habituation for the 45-volt stimulus, significant habituation for the 25-volt stimulus, and significantly greater habituation for the 7-volt stimulus (p <.01 for each comparison; sign test). Thus rate and degree of habituation are inversely related to stimulus strength.

The development of stimulus generalization of habituation is shown in Figure 1E. Single shocks were delivered to two separate branches of the saphenous nerve. The habituating stimulus was given 1/second, and the generalization test stimulus to the other nerve branch was given 1/minute. Data are averages over 3-trial segments at comparable times for the habituating stimulus. Response to the training stimulus habituated to 10% of control level, and the test-stimulus response decreased to about 60% of control level. Amplitudes of the last 10 habituated, generalized, and control responses all differed significantly from each other (habituated versus control, $t = 18.1$, $df = 18$, p <.01; generalized versus control, $t = 2.43$, $df = 18$, p <.05; habituated versus generalized, $t = 7.58$, $df = 18$, p <.01). Thus there is significant but not complete generalization of habituation from one input channel to another.

The effect of "below-zero" habituation on spontaneous recovery is shown in Figure 2A. Stimuli were brief trains to the sural nerve every 5 seconds. In the recovery curves following habituation to "zero," recovery began when the response had reached a stable level. For the recovery curves following habituation below zero, habituation stimuli were continued for 10–20 minutes after the response had reached a stable level. As seen, spontaneous recovery has a longer time course following below-zero habituation. The habituated levels (i.e., zero recovery time) are comparable for both. The two conditions were alternated 22 times in two separate experiments. The mean spontaneous recovery time following habituation to zero was 2.1 minutes, significantly less than the 7-minute mean recovery time following below-zero habituation ($t = 4.45$, $df = 20$, p <.01).

Two examples of dishabituation are shown in Figure 2B. In both cases the response was first habituated to single-shock stimulation of the tibial nerve at 1/3.2 seconds. Weak dishabituation was achieved by increasing the stimulus frequency to 3/second and strong dishabituation by a comparable increase in frequency plus an intensity increase. Habituating stimuli were continued in each series. Data are averages of 8-trial blocks. The weak dishabituating stimulus yielded incomplete recovery, and the strong stimulus resulted in a response of greater than control amplitude. In both cases the response

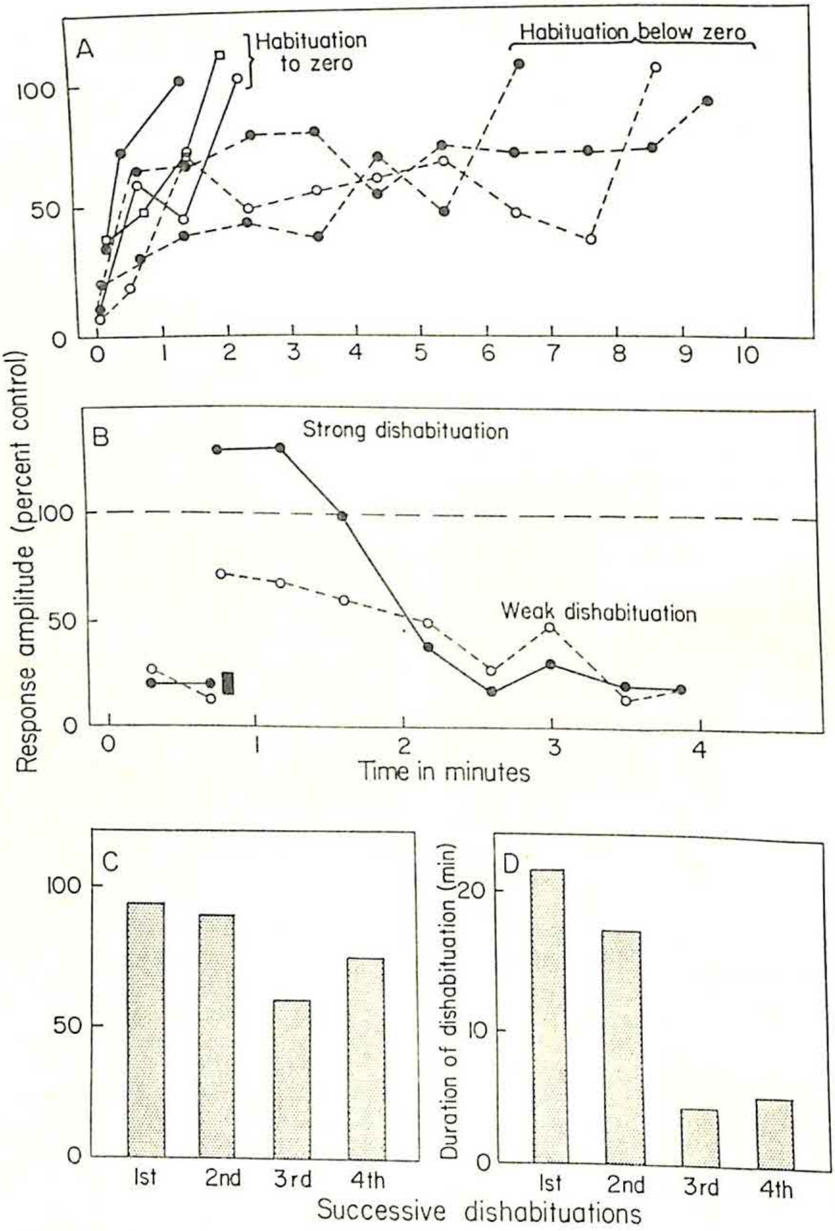

FIG. 2. A: Effect of below-zero habituation on the time course of spontaneous recovery. (Stimuli were brief trains to the sural nerve 1/5 seconds during habituation and 1/minute during the recovery curves shown. Habituation in the to-zero condition was habituation training until the response reached a stable amplitude, and below zero was 10 minutes of additional habituation training after this point had been reached. The to-zero and below-zero conditions were alternated. B: Effects of weak and strong dishabituation on an habituated response. Habituating stimulus was a single shock to the tibial nerve 1/3.2 seconds. Weak dishabituation, bar, was an increase in stimulus frequency to 3/second for 10 seconds, and strong dishabituation was an increase in both frequency and intensity for 10 seconds. Data are averages of 8-trial blocks. C: Habituation of dishabituation. Bars represent the percentage of increase in response amplitude following successive dishabituations delivered when the response was at a stable habituated level. Habituation conditions as in 1A. Dishabituation was a strong pinch of the digits for 5 seconds. D: Same as 2C only duration of the dishabituation effects are plotted.)

returned to habituated levels in about 4 minutes.

Figures 2C and D indicate habituation of dishabituation. Here four successive dishabituating stimuli were given with recovery allowed after each. The habituating stimulus was a brief train of shocks to skin every 10 seconds. Dishabituation was "natural" in this experiment—strong pinching of the digits. There is a reduction both in duration of dishabituation and in amplitude of dishabituated responses following repeated dishabituation.

The data given above illustrate all nine parametric relations characteristic of response habituation in the intact animal. Consequently the flexion reflex of the acute spinal cat exhibits habituation as genuine as that for any response of the intact animal. The time relations for habituation and spontaneous recovery here are within the range of those for responses of the intact animal. It is perhaps somewhat remarkable that a weak stimulus delivered every few seconds for a few minutes can produce a relatively long-lasting but reversible alteration in the behavior of this very much "simplified" nervous system.

Independence of Habituation and Dishabituation

As noted earlier, dishabituation by extraneous stimulation is customarily believed due to disruption of the habituation process, whatever the latter may be. If this were the case, dishabituation ought not to increase the habituated response above its initial control level. However in Figure 2B a strong dishabituatory stimulus was seen to increase the response *well above* the control level. This observation suggests an alternative hypothesis, namely that dishabituation is a separate facilitatory process superimposed upon the habituated system and probably many other response systems as well. Several testable deductions may be made from this hypothesis. First, nonhabituated responses ought to exhibit dishabituation as well as habituated responses. Figure 3A shows this to be the case; the nonhabituated control-level response is increased markedly by a dishabituatory stimulus.

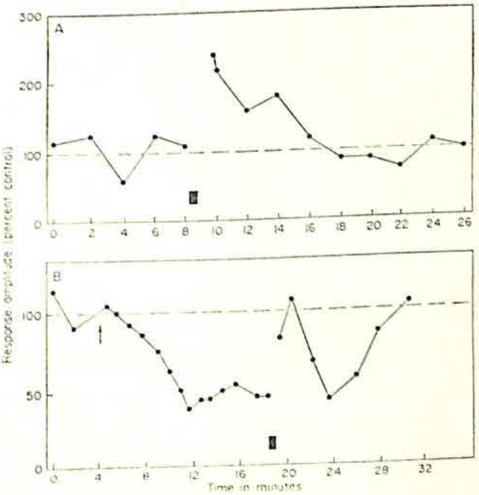

FIG. 3. A: Effect of a strong dishabituating stimulus on an unhabituated (i.e., control-level) response. (Test stimulus was a brief train of shocks to skin every 2 minutes, and dishabituation was a strong shock train delivered elsewhere on the limb for 15 seconds. B: Independence of habituation and dishabituation. Habituation stimulus, arrow to bar, was brief skin-shock train given 1/second and test stimulus given 1/minute. Following the strong, pinch, dishabituatory stimulus, bar, the habituating stimulus was discontinued and a test stimulus delivered 1/minute. Note that the response dishabituates, *then rehabituates* and spontaneously recovers, even though there is no further habituation training.)

A more crucial deduction concerns the role of the habituating stimulus in the decay of the dishabituation effect. If dishabituation is purely a disruption of habituation, a response which is dishabituated following habituation ought not to rehabituate unless additional habituation stimuli are given. Thus in Figure 2B if the habituating stimulus was turned off after dishabituation, the response should remain at the high dishabituated or control level. Alternatively, if dishabituation is a superimposed facilitatory process having an intrinsic decay time, the dishabituated response should gradually decrease toward the habituated level, *even if no habituation training is given subsequent to dishabituation.* This cruical test of the two alternative hypotheses is illustrated in Figure 3B. After habituation to brief trains at 1/second (arrow to bar), the response is dishabituated (bar), and the habituating stimulus withheld and introduced only 1/minute to test response amplitude. As seen, the response amplitude declines (or "rehabituates") almost to the habituated level following dishabituation, *even though it is no longer being habituated.* Subsequently it recovers spontaneously to the control level. This finding has been consistently obtained in more than 20 separate experiments. To demonstrate it clearly, conditions must be such that spontaneous recovery has a long time course relative to dishabituation, otherwise the habituated response will partially recover spontaneously by the time dishabituation effects are dissipated. Since dishabituation is not a disruption of habituation, but rather a separate process having all the characteristics of response sensitization (Hilgard & Marquis, 1940), it will hereafter be termed *sensitization.*

Although the independence of habituation and sensitization has not previously been demonstrated as such, several investigators have suggested the possibility. Humphrey (1933), in his studies of habituation in the snail and turtle, noted that a strong dishabituatory stimulus caused a response of greater amplitude than prehabituated control responses and would also increase control-level responses prior to habituation. His comment deserves quoting: "These results at once suggest . . . that what has been termed 'dishabituation' is simply an increase in excitability, which, under conditions not known, may summate with repeated stimulation [p. 143]." Sharpless and Jasper (1956) suggested that dishabituation of habituated EEG arousal was probably a transient superimposed facilitation. Pavlov (1927), it should be noted, used the same paradigm we employed in Figure 3B to demonstrate that the disinhibitory effect of an extraneous stimulus on a previously extinguished CR was also transient and had the same decay time whether or not the unreinforced conditioned stimulus (CS) was repeatedly applied.

Sensitization thus emerges as a separate and significant process in its own right. A strong stimulus of very short duration can cause an increase in the excitability of the spinal cord lasting for many minutes. Since sensitization is independent of habituation, at least for the spinal flexion reflex, parametric relations Numbers 8 and 9 are not really relevant to habituation as such, but rather describe a separate process. In maintaining the comparability of habituation characteristics for responses of the spinal cord and the intact animal, we are forced to the prediction that the independence of habituation and sensitization will also hold for responses of the intact animal. Future research will determine the extent to which this is true. The habituation of sensitization (characteristic Number 9), incidentally, now becomes simply an additional example of habituation, the response being increased excitability of spinal reflexes by the sensitizing stimulus.

Neuronal Mechanisms of Habituation and Sensitization

Analytic studies of possible neuronal mechanisms involved in habituation and sensitization of the spinal flexion reflex are presented elsewhere (Spencer et al., 1964; also papers in preparation). Results of these experiments will be summarized briefly here. Receptor adaptation was ruled out by electrical stimulation of afferent nerves with monitored neurograms indicating no change in input volleys. Effector fatigue was eliminated as a significant variable by recording activity from efferent nerves and ventral roots rather than muscles. The possible role of feedback from muscle activity was eliminated by nerve and root section and by drug paralysis of neuromuscular junctions, including those between gamma motor neurons and intrafusal muscle fibers. These results indicate that habituation and sensitization occur within the spinal cord. In numerous experiments testing possible habituation of the monosynaptic reflex we have found no evidence of habituation in the sense of progressive decrement in response having a time course comparable to flexion reflex habituation.

The excitability of motor neurons involved in a habituated response can be tested by interposing monosynaptic volleys activating them. In addition, using both ventral-root electrotonus as a population measure, and intracellular microelectrode recording of identified motor-neuron membrane potentials for behavior of neuronal elements, the subliminal responses of motor neurons during habituation and dishabituation were also studied. There was no background alteration in the monosynaptic reflex excitability of motor neurons during habituation, but there was a clear-cut decrease in motor-neuron polysynaptic potentials. Thus whatever processes are responsible for habituation occur prior to motor neurons (i.e., in interneuron pathways) and cannot be due to inhibition acting on the motor neurons. During sensitization there was often, although not invariably, an *increase* in the excitability of motor neurons. These observations further substantiate the behavioral demonstration of the independence of habituation and sensitization.

The possibility that habituation is due solely to altered conduction in afferent terminals of nerve fibers stimulated during training is ruled out by the generalization experiment of Figure 1E. If habituation were due entirely to changes in afferent terminals of the habituated channel, there would be no generalization of habituation to the other channel since completely separate input channels were used. Generalization of habituation did occur, so the decrement must have involved central processes shared by both inputs.

To summarize, it has been proven beyond reasonable doubt that habituation of the spinal flexion reflex is not dependent upon receptor adaptation or effector fatigue and occurs after the afferent input terminals but prior to the output motor neurons in the spinal cord. Consequently the process of habituation must occur in the course of transmission through interneurons. Prosser and Hunter (1936), incidentally, suggested some time ago that this was probably the case. Sensitization, on the other hand, is reflected by increased excitability of both motor and other neurons. Hypotheses and further evidence concerning the processes responsible for habituation and sensitization are developed below.

THEORETICAL ISSUES—BEHAVIORAL

Extinction

The similarity of habituation and experimental extinction was perhaps first emphasized by Humphrey (1930), although Pavlov (1927) had earlier insisted on the basic comparability of unconditioned and conditioned responses. Humphrey (1930, 1933) noted that close parallels exist between the two processes in terms of spontaneous recovery, below-zero effect, and dishabituation. Konorski (1948), in discussing the general properties of plasticity, notes several other parallels including the effect of stimulus frequency and potentiation of habituation upon repeated testing.

The experimental literature on extinction of conditioned responses actually contains many illustrations of the nine parametric relations listed above for habituation (cf. Brogden, 1951; Hull, 1943; Kimble, 1961; Pavlov, 1927). In fact many of these are among the commonly cited characteristics of experimental extinction. To assert that habituation is really extinction does not of course constitute any kind of explanation for either process. By definition extinction presupposes prior conditioning, and the variables of *training* have a potent influence on the characteristics of extinction. Although unspecified prior conditions of learning may modify or shape many "reflex" activities of the intact organism, habituation does imply an unlearned response. However the fact that habituation and extinction appear to show similar parametric features does raise the possibility that they may share common underlying mechanisms. Although many different pieces of the nervous system are clearly involved in extinction and habituation of various responses, we would like to suggest (in agreement with Sharpless, 1964) that common *neuronal* mechanisms exist. Response decrements would then be due to a general property of neuronal activity common to all such systems. At least partial analysis of such neuronal mechanisms seems possible in the spinal animal. Hypothetical neural mechanisms consistent with current knowledge of neurophysiology are developed below.

Stimulus Satiation and Related Constructs

A variety of complex behavioral responses such as spontaneous alternation, exploration, and curiosity have been subsumed under the construct of stimulus satiation (Glanzer, 1953). Glanzer's ingenious theory begins with a five-part postulate describing the properties of stimulus satiation and deduces a number of testable predictions about maze running and exploratory behavior. The majority of such predictions have been verified, particularly where deductions from Hull's postulated reactive inhibition lead to opposite predictions (Dember, 1961; Glanzer, 1953). Glanzer's postulate can be viewed as a *formalized restatement of some of the parametric characteristics of response habituation listed above.* All of Glanzer's deductions may also be made from the earlier established properties of response habituation. His general postulate reads: "Each moment an organism perceives a stimulus object or stimulus objects, A, there develops a quantity of stimulus satiation to A [p. 259]." We may rephrase it as: Each moment an organism responds to a stimulus object or stimulus objects, A, there develops a quantity of *response habituation* to A. Glanzer's deductions follow equally well from this restatement of stimulus satiation in terms of response habituation.

Glanzer's approach has been categorized as a peripheral rather than a central theory (Dember, 1961; Walker, 1958). The above review of neurophysiological studies indicated that there is as yet no entirely satisfactory or consistent evidence for habituation or decrement of evoked responses in primary sensory systems at any level from receptor to cortex. In consequence it would seem misleading to label as *stimulus* satiation a decremental process that *does not appear to exist at any level of neural activity which can be identified with the immediate responses to sensory activation.* Insofar as behavioral responses are concerned, the formal properties of stimulus satiation and response habituation are the same.

Walker (1958) suggested that the term "action decrement," having central locus, be substituted for concepts like stimulus satiation or reactive inhibition. Although he does not relate it to the literature on behavioral habituation, its properties appear to be the same. Dember (1956, 1961) and Dember and Earle (1957) have developed still another theoretical framework for this category of behavior which they term "stimulus change." The organism seeks to obtain the "ideal" amount of "stimulus complexity." Thus in an alternation experiment, the animal alternates choices not because he is satiated or habituated to one stimulus, but because the other attracts him more (Dember, 1961). Dember (1956) reported an experiment designed as a critical test of stimulus-satiation versus stimulus-change hypotheses. Rats were given experience in a T maze (one arm white, one black) and then tested with both arms the same. He argued that satiation theory would predict no difference in test-choice behavior, but stimulus change would predict more choices of arm different in experience and test sessions. Dember's results verified his prediction. This argument would seem justified if stimulus satiation is limited *only* to alley brightness. However, if the total response pattern of running, orienting, turning, and response to brightness is considered, then generalization of response habituation will be greater to arm same than arm different. Thus analysis in terms of response habituation would also predict more choices of arm different than arm same. As yet there seems to be no clear way of differentiating experimentally between the alternative hypotheses of "avoiding" or habituating to one stimulus complex versus "approaching" another stimulus complex. It has been suggested (Brown, 1961) that this distinction may be fundamentally untestable. Again, the formal behavioral properties of Dember's and Earle's constructs do not appear to differ greatly from those listed above.

To summarize, the formal behavioral properties of stimulus satiation, action decrement, and search for ideal stimulus complexity appear similar to the properties of response habituation. They may well be indistinguishable at the behavioral level. The earlier and more neutral term, habituation, viewed as response decrement, might better serve the search for central neural correlates.

Reactive Inhibition

The construct of reactive inhibition (I_r), developed by Hull (1943) from Pavlov's (1927) more general concept of "inhibitory state," has been shown to have considerable predictive utility, particularly in experiments relating to work performance (Hull, 1943; Solomon, 1948). It is unquestionably the most carefully and extensively analyzed construct dealing with response decrement.

The parametric characteristics of habituation listed above are consistent with deductions from reactive inhibition, except for Number 5, indicating that amount of habituation is inversely related to stimulus strength. At first glance this may seem contradictory, in the sense that response strength is often related to stimulus strength. However, I_r has been developed within the framework of performance variables, where the stimulus events cannot easily be specified, and no close relation between stimulus strength and response amplitude holds. In habituation, particularly of spinal reflexes, initial response strength is directly related to stimulus strength. The increased activation of the response by a stronger stimulus could overcome the I_r buildup. Sherrington's analysis of reflex fatigue (see below) elaborates this point.

Glanzer (1953) showed that deductions from I_r in T-maze alternation behavior appeared to lead to the empirically contracted prediction that subjects will alternate turns even if the discriminable stimuli on the two sides are switched each trial. This has been considered a telling argument against I_r (Dember, 1961; Glanzer, 1953). It seems to us that it may be a rather artificial contradiction resulting from the definition of response which was chosen. From the actual response sequence of running down the alley, *orienting toward and approaching one of the stimulus complexes,* and hence turning, Glanzer abstracts turning as *the* response. If we instead emphasize stimulus *orientation and approach* as the more relevant aspect of the response sequence, then application of Hull's I_r analysis leads to exactly the same predictions as does Glanzer's satiation hypothesis. This characterization of the response, incidentally, is comparable to that used by Spence (1960, p. 374) in his analysis of spatial discrimination learning.

To summarize, I_r is not inconsistent with the phenomena of habituation, and may even be viewed as a molar analysis of the role of response habituation in behavior. Experimentally contradicted predictions made from I_r about alternation behavior may be ascribed to the particular response definition used. The hypothetical neuronal model of habituation developed below might well be considered as a possible neurophysiological basis for certain aspects of I_r.

Fatigue

There is no agreed upon definition of fatigue, either in psychology or physiology. We have defined *effector fatigue* as response decrement localized to neuroeffector and effector processes (cf. Merton, 1954) and shown that it cannot account for habituation, at least of the flexion reflex. It is extremely unlikely that effector fatigue has any great importance for habituation generally since all responses that show habituation can be elicited more strongly, more rapidly, and without decrement under other conditions. As far as central fatigue is concerned, arbitrary definitions specifying certain synaptic states could also be used, but have not come into general parlance. Sherrington's (1906, p. 220) molar definition of reflex fatigue is perhaps the most commonly used: ". . . a spinal reflex under continuous excitation or frequent repetition becomes weaker, and may cease altogether." He was very careful to hold firmly to this purely operational definition, which is of course identical to the definition of habituation given above. In like manner, Hull (1943, p. 278) identifies reactive inhibition with fatigue, the latter being defined as ". . . decrement in action evocation potentiality *rather* than an *exhaustion* of the energy . . . [italics ours]."

The inverse relationship between stimulus strength and rate and degree of fatigue or habituation was analyzed very carefully by Sherrington (1906, p. 220):

> In my experience, these spinal reflexes fade out sooner under a weak stimulus than under a strong one. This seeming paradox indicated that even under feeble intensities of stimulation the threshold of the reaction gradually rises, and that it rises above the threshold value of the weaker stimulus before it reaches that of the stronger stimulus.

Many students of habituation have argued against fatigue as the mechanism. Most of these have implied effector fatigue (e.g., Harris, 1943; Prosser & Hunter, 1936). However the more general concept of fatigue has also been questioned on the grounds that dishabituation could not occur if a simple fatigue process were responsible for decrement. Such arguments of course make an implicit assumption about the nature of central fatigue; namely, that it is a process which cannot be broken up rapidly by dishabituation. However when dishabituation is viewed as an independent superimposed process (i.e., sensitization), the argument loses its force. Thus in the flexion reflex, response decrement occurs in interneuronal pathways without any background change in motor-neuron excitability, but sensitization is commonly reflected in transient depolarization of motor neurons as well.

In summary, operational definitions of central fatigue do not differ from definitions of habituation and reactive inhibition. Even if this central process is characterized as having a slow recovery time, dishabituation is not a contradiction if viewed as an independent superimposed sensitization process.

THEORETICAL ISSUES—NEURAL

Stimulus Model Comparison

On the basis of a wide variety of ingenious experiments concerning habituation of the human orienting reflex, Sokolov (1960) has developed a theoretical neural schema. While this theory is somewhat specific to the orienting response, it has occasioned widespread interest (cf. Magoun, 1963; Pribam, 1963) and can be generalized to other types of response habituation. In essence it is a feedback comparator analogue. One system (cerebral cortex) is concerned with the formation of a model of the stimulus-input pattern. This interacts with an amplifying system (reticular formation) which serves to alter response tendency. As the model develops in the cortex, reticular amplification is inhibited. If a dishabituating stimulus is given, the amplification is no longer inhibited, and an increased response occurs. The essential point in the theory is that a model of the habituating stimulus activity develops, against which any other stimulus pattern is compared. Response strength is directly proportional to the degree of discordance between any stimulus input and the model.

The crucial experimental observations forming the basis of this theory concern dishabituation by altered stimulus conditions where intensity is not increased. Thus after habituation to a given tone, a weaker tone or a shorter tone produces dishabituation[5] of the orienting reflex (Sokolov, 1960). Hence a comparison model is necessary. There are few direct tests of Sokolov's theory. Habituation of the orienting reflex does occur, albeit more slowly, in the totally decorticated animal (Lebidinskaia & Rosenthal, 1935), and auditory orientation habituation occurs in animals without auditory cortex (Thompson & Welker, 1963). These observations tend to raise difficulties for the role of the cortex as the model formation system, but the general construct could still be used by invoking primary sensory relays below the thalamocortical level for the model comparater. In like manner it could be generalized to spinal reflexes by assuming a dual mechanism (as yet unidentified) in cord internuncials.

Perhaps the most important point about the theory is that it may not be necessary to assume a model-comparison process to account for dishabituation by reduced stimulus intensity or altered stimulus characteristic. Coding of intensity in neural networks is complex, and there is not always a simple relation between stimulus intensity and neural activity. Thus in the inferior colliculus and auditory cortex, for any given frequency, some neurons fire to weak tones but not to strong tones (Hind, Rose, Davies, Woolsey, Benjamin, Welker, & Thompson, 1960; Rose, Greenwood, Goldberg, & Hind, 1963). Hence a weak tone will activate cells and synapses not activated by the stronger tone. If habituation is simply some sort of decrease in synaptic efficacy, then a larger response would be expected to a weak tone after habituation to a strong tone because different neurons are activated.

---

[5] It should be noted that this is a somewhat special use of the term dishabituation in that the extra stimulus is subsequently used as the test stimulus.

Consequently no sort of comparison against a model is necessary; the process can equally be construed as straightforward decreases in the efficacy of certain pathways, together with differing degrees of pathway overlap resulting from differing stimulus conditions. *Dishabituation by change to a weaker or otherwise altered stimulus then becomes an instance of incomplete stimulus generalization of habituation.* A stimulus having somewhat different central connections than the habituating stimulus yields a larger net response because generalization of habituation to the new stimulus is only partial.

Afferent Neuronal Inhibition

This theory has been elaborated by Hernández-Peón (1955) and has received considerable attention (e.g., Livingston, 1959; Worden & Livingston, 1961). In essence it is assumed that centrifugal fibers from the reticular formation and related structures coursing to various peripheral sensory receptors or nuclei act to inhibit responsiveness at the periphery, thus reducing activity in sensory systems. This inhibition is said to develop during repetitive stimulation and results in evoked response habituation.

There are sound neurophysiological reasons for believing that some types of efferent control systems do exist, although they do not always involve the reticular formation. Two such systems have been described for the auditory system, one acting on the cochlea (Galambos, 1965; Rasmussen, 1946) and one on the cochlear nucleus (Desmedt, 1960; Desmedt & Michelse, 1958). However neither is coextensive with the reticular formation. For the optic system, Granit (1955) has demonstrated both excitatory and inhibitory effects on retinal ganglion cells by stimulation of the midbrain tegmentum. In the somatic system, Hagbarth and Kerr (1954) showed that reticular stimulation could yield both facilitatory and inhibitory effects on the second-order ascending sensory neurons in the spinal cord. Kerr and Hagbarth (1955) found diminution of olfactory-bulb activity when stimulating a variety of nonreticular structures such as prepyriform cortex, amygdaloid nucleus, and olfactory tubercle.

In sum, efferent control systems appear to exist for all sensory modalities studied. Control is sometimes exerted on early central relays rather than on peripheral receptors and does not always involve the reticular formation. These systems could well provide the basis for habituation of sensory evoked primary responses. Unfortunately, as yet there seems to be no entirely satisfactory or consistent evidence for the occurrence of such habituation (except for rhythmic activity of the olfactory bulb, Moulton, 1963). The further possibility that EEG arousal habituation is so mediated is negated by studies showing comparable habituation phenomena for EEG arousal induced by central electrical reticular stimulation (Glickman & Feldman, 1961) where peripheral gating could not have any role, and by Sharpless and Jasper's (1956) finding that primary evoked responses did not habituate during habituation of EEG arousal to auditory stimuli.

Alterations in Synaptic Transmission

In the course of this article it has been suggested at several points that common neuronal mechanisms may account for many varieties of response habituation. Sherrington (1898, 1906), Prosser and Hunter (1936), Sharpless and Jasper (1956), and Kennedy and Mellon (1964), among others, have all raised the possibility that some type of decrease in synaptic transmission may form the basis of habituation. An hypothetical mechanism will be developed here for the spinal flexion reflex, and its possible applicability to other systems considered.

It was proved above that habituation of the flexion reflex in the neurally isolated spinal cord must occur in the transmission through interneurons between incoming afferent fibers and output motor neurons. At least three general classes of neural events may be postulated to yield decreased neuron response tendency: synaptic inhibition, decreased synaptic efficacy, or membrane desensitization.

Synaptic Inhibition. Within the category of synaptic inhibition, two general types have been described, postsynaptic inhibition (PostSI), reflected in hyperpolarizing inhibitory postsynaptic membrane potentials, and presynaptic inhibition (PreSI), which acts by depolarizing presynaptic terminals to yield a smaller excitatory postsynaptic potential (Eccles, 1964). These inhibitory mechanisms have been developed for several types of neurons in the spinal cord and are now being established for certain neurons in other portions of the CNS (Eccles, 1964).

Both types of inhibition could be postulated to act on interneurons mediating the flexion reflex. However, the measured time courses of both these types of inhibition (Eccles, 1964) are much too short to summate with 1/second stimulation and to account for effects lasting several minutes. It could be postulated that complex networks involving tonic or cascading and cumulative PreSI or PostSI are activated by the repeated-stimulus pattern. Time relations are still a problem since Gerard and Forbes (1928) were able to detect some homonymous depression up to .8 second after only a single shock (to mixed nerves in their experiment), and we have found a similar depression with a 1-second interval. The time course of known types of Pre- and PostSI is less than about 500 milliseconds. Evidence from pharmacological studies is also against an inhibitory hypothesis. Strychnine has been shown to attenuate greatly all types of spinal-cord PostSI on which it has been tested (Eccles, 1964). Picrotoxin has been shown to partially block PreSI (Eccles, 1964). We have found that habituation of the flexion reflex is not significantly altered following administration of these two agents in doses sufficient to obtain the reductions in inhibition just noted. Finally, it may be mentioned that electrical indexes of phasic PreSI (such as the cord-dorsum P wave) and phasic PostSI (hyperpolarizing phases of intracellularly and electrotonically recorded motor-neuron responses) themselves show decrement upon repeated activation of cutaneous afferents (Spencer et al., 1964). The N_1 component of the cord-dorsom response, which itself is highly susceptible to presynaptic inhibition, often shows no decrement during the development of habituation, thus providing further evidence against PreSI. To summarize, all existing evidence points away from inhibition as the agent underlying this type of habituation.[6]

Decreased Synaptic Efficacy. In considering alternative explanations, it is helpful to recall that there are at least two other well-known consequences of reflex action which are quite different from inhibition. The first is a species of synaptic depression variously termed "low frequency depression," "defacilitation," "homosynaptic depression," or "subsynaptic depression" which has been studied in the monosynaptic reflex pathway (Curtis & Eccles, 1960; Jefferson & Schlapp, 1953; Lloyd & Wilson, 1957) and also in some of the simplest disynaptic systems (Wilson,

---

[6] However it is to be noted that our experiments were carried out with single-shock stimuli at 1–3-second intervals and that inhibition, particularly the presynaptic variety, may play an important role when higher frequencies are used, as they were in Sherrington's (1899, 1906) studies on fatigue or when intermittent stimulus trains are employed.

1958). Several features of this depression are noteworthy within the context of our discussion: (*a*) It is manifested as a decrease of PSP amplitude due to intermittent low-frequency synaptic activation and is either pre- or subsynaptic rather than postsynaptic in origin; (*b*) its time course is too long to be accounted for by presynaptic inhibition (Eccles, Kostyuk, & Schmidt, 1962); and (*c*) it is detectable at those stimulus intervals required to produce the polysynaptic decrement we have described using single-shock stimuli. The degree of low-frequency depression is greater with increased stimulus frequency, as is the case with habituation. Although Jefferson and Schlapp (1953), recording ventral-root volleys, did report diminished depression with strong stimuli, which would further parallel habituation, Curtis and Eccles (1960), recording motor-neuron excitatory postsynaptic potentials, did not find this effect even with stimuli supramaximal for Group Ia fibers. The depression produced in monosynaptic pathways by low-frequency depression is briefer in duration (less than 20 seconds) than that which we have described. However, it must be remembered that the time course of the depression might be modified in our experiments due to the fact that interneurons characteristically fire in brief trains of discharges, even to single presynaptic volleys, and that this would constitute a distinctly different test pattern which might have a more prolonged recovery period.

The second relevant consequence of reflex action, "after discharge" (Forbes, 1922; Ranson & Hinsey, 1930), is opposite in effect and characteristic of polysynaptic systems. This is manifested as a postsynaptic depolarization and prolonged discharge of inter and motor neurons (Fuortes, 1954; Hunt & Kuno, 1959), is generally more prolonged with stronger stimuli, and may build up with repeated stimuli (Frank & Fuortes, 1956). Preliminary evidence we have obtained from extracellular interneuron recordings indicates that such afterdischarges are present and build up with stimulus repetition during the processes described.

We would therefore suggest, as a working hypothesis, that the amplitude of repeated polysynaptic test responses depends upon the balance between low-frequency depression and afterdischarge existing at the time each stimulus is presented. Viewed in this way, the frequency sensitivity of the decrement we have described would be related to the well-known frequency sensitivity of low-frequency depression. Stimulus generalization would depend upon depression of interneuron synapses shared by the two input channels. The greater response stability seen at the strong stimulus strengths might reflect a greater preponderance of afterdischarge onto interneurons relaying the reflex. This would increase in duration with strong stimuli and would exist during the presentation of successive stimuli, thereby facilitating the postsynaptic interneuron responses to them, in a sense compensating for the low-frequency depression. Moreover, if the phenomenon of sensitization (i.e., response restoration following a strong extrastimulus) be considered a special case of very prolonged afterdischarge due to a strong stimulus, then many of the puzzling features of this phenomenon could be more readily understood. Wilson's (1955) postulated mechanism of polysynaptic posttetanic potentiation (PTP) is an alternative possibility for sensitization, as is the recently described phenomenon of "presynaptic facilitation" (Kandel & Tauc, 1964; Mendell & Wall, 1965), since there are certain obvious similarities to the process of response restoration we have described.

Membrane Desensitization. We noted earlier that some forms of response habituation, even of spinal reflexes, may last for days or weeks. In the experiment by Kozak et al. (1962), for example, a crossed scratch reflex of the chronic spinal kitten, elicited by rubbing an area of skin for 10 minutes daily, exhibited marked habituation, requiring a period of 8 days to recover. The synaptic processes discussed above, namely inhibition and decreased efficacy due to the balance of low-frequency depression and afterdischarge, would both seem to have too short a time course, per se, to account for such long-lasting habituation. Sharpless (1964) has suggested that a type of membrane desensitization process, involving decreased responsiveness of the postsynaptic membranes of nerve cells due to repeated applications of the transmitter substances, may be involved in long-term habituation. In hypersensitized denervated smooth muscles and glands, for example, administration of appropriate excitatory agents such as epinephrine several times daily produces a desensitization effect that may last for several days (Emmelin & Muren, 1952; Simeone, 1938). This type of hypothesis may be particularly appropriate to the chronic spinal preparation in that spinal reflexes become markedly exaggerated or hypersensitized, as do responses of dennervated smooth muscles and glands. Two different classes or components of responses have been distinguished on the basis of habituation and spontaneous-recovery time courses. Sokolov (1955) noted that the "orienting reflex" exhibited both rapidly habituating and slowly habituating components. Sharpless and Jasper (1956) differentiated between a "phasic" EEG arousal that occurs immediately after a stimulus, is of short duration, and is relatively resistant to habituation and a "tonic" component that habituates much more rapidly. Lesions of the brachia of the inferior colliculus appeared to abolish the phasic component (to auditory stimuli) without eliminating the tonic component. Thompson and Welker (1963) found that ablation of auditory cortex prevented the development of long-term intersession habituation of orienting reflexes to sounds, but did not alter short-term intrasession habituation.

It may well be necessary to postulate at least two different types of neuronal processes to account for short-term and long-term habituation. Our suggested mechanisms of low-frequency depression and afterdischarge are perhaps appropriate to short-term effects and Sharpless' "membrane desensitization" to long-term effects. Alternatively, all these processes may interact to a greater or lesser extent in all instances of habituation. The fact that, except for time course, both short-term and long-term habituation exhibit the same parametric features may imply some identity of underlying processes.

CONCLUSION

In summary, we would suggest that, at least for the spinal flexion reflex as we have studied it, the process of habituation may represent the cumulative effect of a *polysynaptic low-frequency depression* and that dishabituation is a type of *superimposed sensitization*, possibly resulting from facilitatory afterdischarge.

In amplifying this scheme, one might postulate, on the basis of other evidence (Frank & Fuortes, 1956), that rather long iterative trains of interneuron discharges are produced by each stimulus, and that the long duration of the changes noted depends on this feature of the test response. If one provisionally adopts this point of view, several aspects of the literature on habituation reviewed here may be seen in somewhat different perspective. The absence or poor development of evoked response habituation in primary sensory systems would be consistent with the fact that the interneurons exhibiting very prolonged discharge patterns to single stimuli are not typical of the long relays from receptors to cortex in these systems. However such iterative firing is more characteristic of neurons in the system mediating EEG arousal; the reticular formation (Moruzzi, 1954), which does exhibit habituation (Bell, Sierra, Buendia, & Segundo, 1964; Sharpless & Jasper, 1956). In like manner, such interneurons are undoubtedly present in all functional systems of the intact organism participating in behavioral response habituation.

Other things equal, complex responses tend to habituate more rapidly than simpler responses. If degree of response complexity is reflected in increased numbers of interneurons interposed between stimulus and response, and in an increase in the duration of their repetitive discharges, these factors themselves might account for more rapid habituation. The effects of CNS lesions on behavioral response habituation may also be reconsidered in this way: Effective lesions remove many interneurons normally involved in complex responses and hence could lead to slower response habituation.

Many of these extrapolations from data obtained in the study of response decrement in the spinal cord to habituation as it appears in the more intact organism are obviously speculative. They do have the virtue, however, of indicating the directions which cellular theories of habituation could take. Quantitative theoretical models would clearly be of great help in making such cellular explanations useful alternatives to the more molar theories commonly applied to behavior of the intact organism.

REFERENCES

ADRIAN, E. D. The electrical activity of the mammalian olfactory bulb. *Electroencephalography and Clinical Neurophysiology,* 1950, **2**, 377–388.

AFELT, Z. Variability of reflexes in chronic spinal frogs. In E. Gutmann & P. Hnik (Eds.), *Proceedings of the conference on central and peripheral mechanisms of motor functions.* Prague, Czechoslovakia: Czechoslovak Academy of Sciences, 1963. Pp. 37–41.

AFFANNI, J., MANCIA, M., & MARCHIAFAVA, P. L. Role of the pupil in changes in evoked responses along the visual pathways. *Archives Italiennes de biologie,* 1962, **100**, 287–296.

ALTMAN, I. A. Electrophysiological examination of different parts of the auditory system in the cat during sustained rhythmical stimulation. *Sechenov Physiological Journal of the USSR,* 1960, **46**, 617–629.

APELBAUM, J., SILVA, E. E., FRICK, O., & SEGUNDO, J. P. Specificity and biasing of arousal reaction habituation. *Electroencephalography and Clinical Neurophysiology,* 1960, **12**, 829–839.

BARD, P., & MACHT, M. B. The behavior of chronically decerebrate cats. In G. E. W. Wolstenholme & C. M. O'Connor (Eds.), *Neurological basis of behavior.* London: Churchill, 1958. Pp. 55–71.

BARTLEY, S. H., & CHUTE, E. *Fatigue and impairment in man.* New York: McGraw-Hill, 1947.

BELL, C., SIERRA, G., BUENDIA, N., & SEGUNDO, J. P. Sensory properties of neurons in the mesencephalic reticular formation. *Journal of Neurophysiology,* 1964, **27**, 961–987.

BERLYNE, D. E. Novelty and curiosity as determinants of exploratory behavior. *British Journal of Psychology,* 1950, **41**, 68–80.

BERLYNE, D. E. *Conflict, arousal, and curiosity,* New York: McGraw-Hill, 1960.

BINDRA, D. Stimulus change, reactions to novelty, and response decrement. *Psychological Review,* 1959, **66**, 96–103.

BROGDEN, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 568–612.

BROWN, J. S. *The motivation of behavior.* New York: McGraw-Hill, 1961.

BUCHWALD, J. S., HALAS, E. S., & SCHRAMM, S. Progressive changes in unit responses to a repeated stimulus. *Federation Proceedings,* 1964, **23**, 919.

BUCHWALD, J. S., HALAS, E. S., & SCHRAMM, S. Progressive changes in efferent unit responses to repeated cutaneous stimulation in spinal cats. *Journal of Neurophysiology,* 1965, **28**, 200–215.

BUTLER, R. A., & HARLOW, H. F. Persistence of visual exploration in monkeys. *Journal of Comparative and Physiological Psychology,* 1954, **47**, 258–263.

CASPERS, H., LERCHE, E., & GREUTER, H. Adaption-serscheinungen der akustich ausgelösten Wechreaktion bei Reizung mit definierten Tonimpulsen. *Pflüger's Archives,* 1958, **267**, 128-141.

CAVAGGIONI, A., GIANNELLI, G., & SANTIBAÑEZ, H. G. Effects of repetitive photic stimulation on responses evoked in the lateral geniculate body and the visual cortex. *Archives Italiennes de biologie,* 1959, **97**, 226-275.

COOMBS, C. H. Adaptation of the galvanic response to auditory stimuli. *Journal of Experimental Psychology,* 1938, **22**, 244-268.

CRAMPTON, G. H., & SCHWAM, W. J. Effects of arousal reaction on nystagmus habituation in the cat. *American Journal of Physiology,* 1961, **200**, 29-33.

CURTIS, D. R., & ECCLES, J. C. Synaptic action during and after repetitive stimulation. *Journal of Physiology,* 1960, **150**, 374-398.

DAVIS, R. C. Modification of the galvanic reflex by daily repetition of a stimulus. *Journal of Experimental Psychology,* 1934, **17**, 504-535.

DAVIS, R. C., BUCHWALD, A. M., & FRANKMANN, R. W. Autonomic and muscular responses, and their relation to simple stimuli. *Psychological Monographs,* 1955, **69**(20, Whole No. 405).

DEMBER, W. N. Response by the rat to environmental change. *Journal of Comparative and Physiological Psychology,* 1956, **49**, 93-95.

DEMBER, W. N. Alternation behavior. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience.* Homewood, Ill.: Dorsey, 1961. Pp. 227-252.

DEMBER, W. N., & EARL, R. W. Analysis of exploratory, manipulatory, and curiosity behavior. *Psychological Review,* 1957, **64**, 91-96.

DESMEDT, J. E. Neurophysiological mechanisms controlling acoustic input. In G. L. Rasmussen & W. Windle (Eds.), *Neural mechanisms of the auditory and vestibular systems.* Springfield, Ill.: Charles C Thomas, 1960. Pp. 152-164.

DESMEDT, J. E., & MICHELSE, K. Suppression of acoustic input by thalamic stimulation. *Proceedings of the Society for Experimental Biology and Medicine,* 1958, **99**, 772-775.

DEUTSCH, J. A. Higher nervous function: The physiological bases of memory. *Annal Review of Physiology,* 1962, **24**, 259-286.

DYKMAN, R. A., REESE, W. G., GALBRECHT, C. R., & THOMASSON, P. J. Psychophysiological reactions to novel stimuli: Measurement, adaptation, and relationship of psychological variables in the normal human. *Annals of the New York Academy of Sciences,* 1959, **79**, 43-107.

ECCLES, J. C. *The physiology of synapses.* New York: Academic Press, 1964.

ECCLES, J. C., KOSTYUK, P. G., & SCHMIDT, R. F. Presynaptic inhibition of the central actions of flexor reflex afferents. *Journal of Physiology* 1962, **161**, 258-281.

EMMELIN, N., & MUREN, A. The sensitivity of submaxillary glands to chemical agents studied in cats under various conditions over long periods. *Acta Physiologica Scandinavica,* 1952, **26**, 221-231.

ENGEN, T., & LIPSITT, L. P. Decrement and recovery of responses to olfactory stimuli in the human neonate. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 312-316.

FERNANDEZ-GUARDIOLA, A., ROLDÁN, E., FANJUL, L., & CASTELLS, C. Role of the pupillary mechanism in the process of habituation of the visual pathways. *Electroencephalography and Clinical Neurophysiology,* 1961, **13**, 564-576.

FORBES, A. The interpretation of spinal reflexes in terms of present knowledge of nerve conduction. *Physiological Review,* 1922, **2**, 361-414.

FRANK, K., & FUORTES, M. G. F. Unitary activity of spinal interneurons of cats. *Journal of Physiology,* 1956, **131**, 424-436.

FRENCH, G. M. Locomotor effects of regional ablations of frontal cortex in rhesus monkeys. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 18-24.

FRENCH, G. M., & HARLOW, H. F. Locomotor reaction decrement in normal and brain-damaged monkeys. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 496-501.

FUORTES, M. G. F. Activity of single motorneurons during some reflex reactions of mammals. *Comptes-Rendus 5th International Neurological Congress* (Lisbon), 1954, **4**, 36-51.

GALAMBOS, R. Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. *Journal of Neurophysiology,* 1956, **19**, 424-437.

GALAMBOS, R., SHEATZ, G., & VERNIER, V. G. Electrophysiological correlates of a conditioned response in cats. *Science,* 1956, **123**, 331-332.

GARCIA-AUSTT, E. Influence of the states of awareness upon sensory evoked potentials. *Electroencephalography and Clinical Neurophysiology Suppl. 24,* 1963, 76-89.

GERARD, R. W., & FORBES, A. "Fatigue" of the flexion reflex. *American Journal of Physiology,* 1928, **86**, 186-205.

Gershuni, G. V., Kozhevnikov, V. A., Maruseva, A. M., Avakyan, R. V., Radionova, E. Z., Altman, J. A., & Soroko, V. I. Modifications in electrical responses of the auditory system in different states of the higher nervous system. *Electroencephalography and Clinical Neurophysiology Suppl. 13,* 1960, 115–124.

Glanzer, M. Stimulus satiation: An explanation of spontaneous alternation and related phenomena. *Psychological Review,* 1953, **60**, 257–268.

Glaser, E. M., & Griffin, J. P. Influence of the cerebral cortex on habituation. *Journal of Physiology,* 1962, **160**, 429–445.

Glickman, S. E. Deficits in avoidance learning produced by stimulation of the ascending reticular formation. *Canadian Journal of Psychology,* 1958, **12**, 97–102.

Glickman, S. E., & Feldman, S. M. Habituation of the arousal response to direct stimulation of the brainstem. *Electroencephalography and Clinical Neurophysiology,* 1961, **13**, 703–709.

Granit, R. Centrifugal and antidromic effects on ganglion cells of retina. *Journal of Neurophysiology,* 1955, **18**, 388–411.

Green, J. D., & Arduini, A. A. Hippocampal electrical activity in arousal. *Journal of Neurophysiology,* 1954, **17**, 533–557.

Griffin, J. P. The role of the frontal areas of the cortex upon habituation in man. *Clinical Science,* 1963, **24**, 127–134.

Griffith, C. R. The effect upon the white rat of continued bodily rotation. *American Naturalist,* 1920, **54**, 524–534.

Guzmán-Flores, C., Alcaraz, M., & Harmony, T. Role of the intrinsic ear muscles in the process of acoustic habituation. *Boletín del Instituto de estudios médicos y biológicos, Universidad nacional de México,* 1960, **18**, 135–149.

Hagamen, W. G., Lance, E. M., & Ungewitter, L. H. Increased responsiveness to stimuli following lesions of the forebrain. *Anatomical Record,* 1959, **133**, 387–388.

Hagbarth, K. E., & Finer, B. L. The plasticity of human withdrawal reflexes to noxious skin stimuli in lower limbs. In G. Moruzzi, A. Fessard, & H. H. Jasper (Eds.), *Progress in brain research.* Vol. 1. *Brain mechanisms.* New York: Elsevier, 1963. Pp. 65–81.

Hagbarth, K. E., & Kerr, D. I. B. Central influences on spinal afferent conduction. *Journal of Neurophysiology,* 1954, **17**, 295–307.

Hagbarth, K. E., & Kugelberg, E. Plasticity of the human abdominal skin reflex. *Brain,* 1958, **81**, 305–318.

Harris, J. D. Habituatory response decrement in the intact organism. *Psychological Bulletin,* 1943, **40**, 385–422.

Hernández-Peón, R. Central mechanisms controlling conduction along central sensory pathways. *Acta neurologica latinoamericana,* 1955, **1**, 256–264.

Hernández-Peón, R. Neurophysiological correlates of habituation and other manifestations of plastic inhibition. *Electroencephalography and Clinical Neurophysiology Suppl. 13,* 1960, 101–114.

Hernández-Peón, R., Guzmán-Flores, C., Alcaraz, M., & Fernandez-Guardiola, A. Habituation in the visual pathway. *Acta neurologica latinoamericana,* 1958, **4**, 121–129.

Hernández-Peón, R., Jouvet, M., & Scherrer, H. Auditory potentials at cochlear nucleus during acoustic habituation. *Acta neurologica latinoamericana,* 1957, **3**, 144–156.

Hernández-Peón, R., Lavin, A., Alcocer-Cuarón, C., & Marcelin, J. P. Electrical activity of the olfactory bulb during wakefulness and sleep. *Electroencephalography and Clinical Neurophysiology,* 1960, **12**, 41–58.

Hernández-Peón, R., & Scherrer, H. Habituation to acoustic stimuli in cochlear nucleus. *Federation Proceedings,* 1955, **14**, 71.

Hernández-Peón, R., Scherrer, H., & Velasco, M. Central influences on afferent conduction in the somatic and visual pathways. *Acta neurologica latinoamericana,* 1956, **2**, 8–22.

Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century-Crofts, 1940.

Hind, J. E., Rose, J. E., Davies, P. W., Woolsey, C. N., Benjamin, R. M., Welker, W. I., & Thompson, R. F. Unit activity in the auditory cortex. In G. L. Rasmussen & W. Windle (Eds.), *Neural mechanisms of the auditory and vestibular systems.* Springfield, Ill.: Charles C Thomas, 1960. Pp. 201–210.

Hinde, R. A. Changes in responsiveness to a constant stimulus. *British Journal of Animal Behavior,* 1954, **2**, 41–55.

Hugelin, A., Dumont, S., & Paillas, N. Tympanic muscles and control of auditory input during arousal. *Science,* 1960, **131**, 1371–1372.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Humphrey, G. Extinction and negative adaptation. *Psychological Review,* 1930, **37**, 361–363.

Humphrey, G. *The nature of learning.* New York: Harcourt, Brace, 1933.

Hunt, C. C., & Kuno, M. Properties of spinal interneurons. *Journal of Physiology,* 1959, **147**, 346–363.

Huttenlocher, P. R. Effects of state of arousal on click responses in the mesencephalic reticular formation. *Electroencephalography and Clinical Neurophysiology,* 1960, 12, 819–827.

James, W. *Principles of psychology.* New York: Holt, 1890.

Jefferson, A., & Schlapp, W. Some effects of repetitive stimulation of afferents on reflex conduction. In J. L. Malcolm & J. A. B. Gray (Eds.), *The spinal cord.* London: Churchill, 1953. Pp. 99–119.

Jouvet, M., & Hernández-Peón, R. Mécanismes neurophysiologiques concernant l'habituation, l'attention, et le conditionnement. *Electroencephalography and Clinical Neurophysiology Suppl. 6,* 1957, 39–49.

Kandel, E. R., & Tauc, L. Mechanisms of prolonged heterosynaptic facilitation. *Nature,* 1964, **202**, 145.

Kennedy, D., & Mellon, D., Jr. Synaptic activation and receptive fields in crayfish interneurons. *Comparative Biochemistry and Physiology,* 1964, **13**, 275–300.

Kerr, D. I. B., & Hagbarth, K. E. An investigation of olfactory centrifugal fiber system. *Journal of Neurophysiology,* 1955, **18**, 362–374.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Knott, J. R., & Henry, C. E. The conditioning of the blocking of the alpha rhythm of the human electroencephalogram. *Journal of Experimental Psychology,* 1941, **28**, 134–144.

Konorski, J. *Conditioned reflexes and neuron organization.* New York: Cambridge Univer. Press., 1948.

Kozak, W., MacFarlane, W. W., & Westerman, R. Longlasting reversible changes in the reflex responses of chronic spinal cats to touch, heat and cold. *Nature,* 1962, **193**, 171–173.

Krechevsky, I. Brain mechanisms and variability: I. Variability within a means-end-readiness. *Journal of Comparative Psychology,* 1937, **23**, 121–138. (a)

Krechevsky, I. Brain mechanisms and variability: II. Variability where no learning is involved. *Journal of Comparative Psychology,* 1937, **23**, 139–163. (b)

Krechevsky, I. Brain mechanisms and variability: III. Limitations of the effect of cortical injury upon variability. *Journal of Comparative Psychology,* 1937, **23**, 351–364. (c)

Lebedinskaia, S. I., & Rosenthal, J. S. Reactions of a log after removal of the cerebral hemispheres. *Brain,* 1935, **58**, 412–419.

Lehner, G. F. J. A study of the extinction of unconditioned reflexes. *Journal of Experimental Psychology,* 1941, **29**, 435–456.

Lethlean, A. K. Habituation in the rat spinal cord—a triggered process. *Federation Proceedings,* 1965, **24**, 517.

Livingston, R. B. Central control of receptors and sensory transmission. In J. Field (Ed.), *Handbook of physiology: Section 1. Neurophysiology.* Vol. 1. Washington, D. C.: American Physiological Society, 1959. Pp. 741–760.

Lloyd, D. P. C., & Wilson, V. J. Reflex depression in rhythmically active monosynaptic reflex pathways. *Journal of General Physiology,* 1957, **40**, 409–426.

Lombard, W. P. The variations of the normal knee jerk and their relation to the activity of the central nervous system. *American Journal of Physiology,* 1887, **1**, 5–71.

Magoun, H. W. Central neural inhibition. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1963.* Lincoln: Univer. Nebraska Press, 1963. Pp. 161–196.

Mancia, M., Meulders, M., & Santibañez-H., G. Changes of photically evoked potentials in the visual pathway of the cerveau isolé cat. *Archives Italiennes de biologie,* 1959, **97**, 378–398.

Marsh, J. T., McCarthy, D. A., Sheatz, G., & Galambos, R. Amplitude changes in evoked auditory potentials during habituation and conditioning. *Electroencephalography and Clinical Neurophysiology,* 1961, **13**, 224–234.

Marsh, J. T., & Worden, F. G. Auditory potentials during acoustic habituation: Cochlear nucleus, cerebellum and auditory cortex. *Electroencephalography and Clinical Neurophysiology,* 1964, **17**, 685–692.

Marsh, J. T., Worden, F. G., & Hicks, L. Some effects of room acoustics on evoked auditory potentials. *Science,* 1962, **137**, 281–282.

Martin, I. Adaptation. *Psychological Bulletin,* 1964, **61**, 35–44.

Melzack, R. On the survival of Mallard ducks after "habituation" to the hawk-shaped figure. *Behavior,* 1961, **17**, 1–16.

Mendell, L. M., & Wall, P. D. Presynaptic hyperpolarization: A role for fine afferent fibers. *Journal of Physiology,* 1965, **172**, 274–294.

Merton, P. A. Voluntary strength and fatigue. *Journal of Physiology,* 1954, **123**, 553–564.

Montgomery, K. C. Exploratory behavior as a function of "similarity" of stimulus situations. *Journal of Comparative and*

Physiological Psychology, 1953, **46**, 129–133.

MORGAN, C. T., & WOOD, W. M. Cortical localization of symbolic processes in the rat: II. Effect of cortical lesions upon delayed alternation in the rat. *Journal of Neurophysiology,* 1943, **6**, 173–180.

MORRELL, F. Electrophysiological contributions to the neural basis of learning. *Physiological Review,* 1961, **41**, 443–494.

MORUZZI, G. The ascending reticular activating system and wakefulness. In J. F. Delafresnaye (Ed.), *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954. Pp. 1–20.

MOULTON, D. G. Electrical activity in the olfactory system of rabbits with indwelling electrodes. In Y. Zotterman (Ed.), *Olfaction and taste.* New York: MacMillan, 1963. Pp. 71–84.

MOUSHEGAN, G., RUPERT, A., MARSH, J., & GALAMBOS, R. Evoked cortical potentials in absence of middle ear muscles. *Science,* 1961, **133**, 582.

MOWRER, O. H. The modification of vestibular nystagmus by means of repeated elicitation. *Comparative Psychology Monographs,* 1934, **9**(45).

NAQUET, R., REGIS, H., FISCHER-WILLIAMS, M., & FERNANDEZ-GUARDIOLA, A. Variations in the responses evoked by light along the specific pathways. *Brain,* 1960, **83**, 52–56.

NESMEIANOVA, T. N. The inhibition of the motor reflex in spinal dogs under conditions of chronic experimentation. *Sechenov Journal of Physiology of the USSR,* 1957, **42**, 281–288.

PALESTINI, M., DAVIDOVICH, A., & HERNÁNDEZ-PEÓN, R. Functional significance of centrifugal influences upon the retina. *Acta neurologica latinoamericana,* 1959, **5**, 113–131.

PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford, 1927.

PORTER, J. M. Adaptation of the galvanic skin response. *Journal of Experimental Psychology,* 1938, 23, 553–557.

PRIBRAM, K. H. Reinforcement revisited: A structural view. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1963.* Lincoln: Univer. Nebraska Press, 1963. Pp. 113–160.

PROSSER, C. L., & HUNTER, W. S. The extinction of startle responses and spinal reflexes in the white rat. *American Journal of Physiology,* 1936, **117**, 609–618.

RANSON, S. W., & HINSEY, J. C. Reflexes in the hind limbs of cats after transsection of the spinal cord at various levels. *American Journal of Physiology,* 1930, **94**, 471–495.

RASMUSSEN, G. L. The olivary peduncle and other fiber projections of the superior olivary complex. *Journal of Comparative Neurology,* 1946, **84**, 141–219.

RHINEBERGER, M., & JASPER, H. H. Electrical activity of the cerebral cortex in the unanesthetized cat. *American Journal of Physiology,* 1937, **119**, 186–196.

ROBERTS, W. W., DEMBER, W. N., & BRODWICK, M. Alternation and exploration in rats with hippocampal lesions. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 695–700.

ROBINSON, E. S. Work of the integrated organism. In C. Murchison (Ed.), *A handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934. Pp. 571–650.

ROBINSON, J., & GANTT, W. H. The orienting reflex (questioning reaction): Cardiac, respiratory, salivary, and motor components. *Bulletin of the Johns Hopkins Hospital,* 1947, **80**, 231–253.

RODGERS, W. L., MELZACK, R., & SEGAL, J. R. "Tail flip response" in goldfish. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 917–923.

ROIG, J. A., & SOMMER-SMITH, J. A. Atenuación específica del sobresalto provacado par excitacion nerviosa central. *Anales de la Facultad de medicina, Universidad de Montevideo,* 1959, **44**, 410–413.

ROSE, J. E., GREENWOOD, D. D., GOLDBERG, J. M., & HIND, J. E. Some discharge characteristics of single neurons in the inferior colliculus of the cat: I. Tonotopical organization, relation of spike-counts to tone intensity, and firing patterns of single elements. *Journal of Neurophysiology,* 1963, **26**, 294–320.

SANTIBAÑEZ, G., TROUCHE, E., & ALBE-FESSARD, D. Étude de l'evolution au cours du temps de l'amplitude des activités corticales et thalamiques évoqueés par des stimuli somatique répétés. *Journal de Physiologie,* 1963, **55**, 335.

SCHOLANDER, T. Habituation of autonomic response elements under two conditions of alterness. *Acta physiologica scandinavica,* 1960, **50**, 259–268.

SCHWARTZBAUM, J. S., WILSON, W. A., JR., & MORRISSETTE, J. R. The effects of amygdalectomy on locomotor activity in monkeys. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 334–336.

SEGUNDO, J. P., ROIG, J. A., & SOMMER-SMITH, J. A. Conditioning of reticular formation stimulation effects. *Electroencephalography and Clinical Neurophysiology,* 1959, **11**, 471–484.

Sharpless, S. K. Reorganization of function in the nervous system—use and disuse. *Annual Review of Physiology,* 1964, **26**, 357–388.

Sharpless, S., & Jasper, H. Habituation of the arousal reaction. *Brain,* 1956, **79**, 655–682.

Shaw, J. A., & Thompson, R. F. Dependence of evoked cortical association responses on behavioral variables. *Psychonomic Science,* 1964, **1**, 153–154.

Sherrington, C. S. Experiments in examination of the peripheral distribution of the fibers of the posterior roots of some spinal nerves. *Royal Society of London,* 1898, **190**, 45–186.

Sherrington, C. S. *The integrative action of the nervous system.* New Haven: Yale Univer. Press, 1906.

Shurrager, P. S., & Culler, E. Conditioned extinction of a reflex in the spinal dog. *Journal of Experimental Psychology,* 1941, **28**, 287–303.

Shurrager, P. S., & Shurrager, H. C. Converting a spinal CR into a reflex. *Journal of Experimental Psychology,* 1941, **29**, 217–224.

Simeone, F. A. Responsiveness of denervated nictitating membrane. *American Journal of Physiology,* 1938, **122**, 650–658.

Sokolov, E. N. The higher nervous activity and the problem of perception. *Proceedings of the XIX International Congress of Psychology,* Montreal, 1955.

Sokolov, E. N. Neuronal models and the orienting influence. In M. A. B. Brazier (Ed.), *The central nervous system and behavior: III.* New York: Macy Foundation, 1960.

Solomon, R. L. The influence of work on behavior. *Psychological Bulletin,* 1948, **45**, 1–40.

Spence, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Spencer, W. A., Thompson, R. F., & Neilson, D. R. Analysis of polysynaptic reflex response decrement in the acute spinal cat. *Physiologist,* 1964, **7**, 262.

Symmes, D. Anxiety reduction and novelty as goals of visual exploration by monkeys. *Journal of Genetic Psychology,* 1959, **94**, 181–198.

Symmes, D. Effect of cortical ablations on visual exploration by monkeys. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 757–763.

Thompson, R. F., & Shaw, J. A. Behavioral correlates of evoked activity recorded from association areas of the cerebral cortex. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 329–339.

Thompson, R. F., & Welker, W. I. Role of auditory cortex in reflex head orientation by cats to auditory stimuli. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 996–1002.

Thorpe, W. H. *Learning and instinct in animals.* London: Methuen, 1956.

Voronin, L. G., & Sokolov, E. N. Cortical mechanisms of the orienting reflex and its relation to the conditioned reflex. *Electroencephalography and Clinical Neurophysiology Suppl. 13,* 1960, 335–346.

Walker, E. L. Action decrement and its relation to learning. *Psychological Review.* 1958, **65**, 129–142.

Webster, W. R., Dunlop, C. W., Simons, L. A., & Aitkin, L. M. Auditory habituation: A test of a centrifugal and a peripheral theory. *Science,* 1965, **148**, 654–656.

Welker, W. I. Variability of play and exploratory behavior in chimpanzees. *Journal of Comparative and Physiological Psychology,* 1956, **49**, 181–185.

Welker, W. I. An analysis of exploratory and play behavior in animals. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience.* Homewood, Ill.: Dorsey, 1961. Pp. 175–226.

Wendt, G. R. Vestibular functions. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1191–1223.

Wilson, V. J. Post-tetanic potentiation of polysynaptic reflexes of the spinal cord. *Journal of Genetic Physiology,* 1955, **39**, 197–206.

Wilson, V. J. Early post-tetanic potentiation and low frequency depression of some group I reflex actions. *Journal of Genetic Physiology,* 1958, **41**, 1005–1018.

Worden, F. G., & Livingston, R. B. Brainstem reticular formation. In D. E. Sheer (Ed.), *Electrical stimulation of the brain.* Austin: Univer. Texas Press, 1961. Pp. 263–276.

Worden, F. G., & Marsh, J. T. Amplitude changes of auditory potentials evoked at cochlear nucleus during acoustic habituation. *Electroencephalography and Clinical Neurophysiology,* 1963, **15**, 866–881.

(Early publication received April 5, 1965)

Psychological Review
1966, Vol. 73, No. 1, 44–58

SIGNAL-DETECTABILITY THEORY OF RECOGNITION-MEMORY PERFORMANCE[1]

THEODORE E. PARKS[2]

University of Wisconsin

Psychophysical signal-detection theory is applied to recognition-memory performance. Old items in the test list are considered to be analogous to "signals" while new items are analogous to "noise only." The resulting fundamental assumptions describe the covert responses which mediate recognition-memory performance as varying continuously in strength. Covert responses to both old and new items are normally distributed with the distance between distributions representing learning and retention. The overt response to an item depends on whether or not S's covert response exceeds an arbitrary criterion. The available evidence suggests that S's criterion will be set such that he will most probably choose a number of items approximately equal to the number of old items in the test.

A test of recognition memory typically involves presenting the subject with a list composed of old items (i.e., items that were in an inspection list presented previously) and new items. He is required to examine each item of the test list and to indicate whether or not it was a member of the inspection list. Two measures of performance are taken: (*a*) the proportion of the old items in the test list which the subject indicates are old and (*b*) the proportion of new items which the subject indicates (incorrectly) are old.

Customarily, these two scores have been combined in a "correction for guessing" which was thought to yield a single "true" measure of memory. Underlying this correction is the assumption that recognition memory is a sort of all-or-none process; that is, that correct recognitions consist of items recognized together with items chosen purely through random guessing. On the contrary, the present paper presents evidence that an item is not either recognized or not recognized, but, rather, the covert events which mediate overt recognition behavior vary in strength along a continuum.

Under this assumption, the subject is faced with a complex decision problem. The present theory attempts to describe this problem and the subject's solution in detail and in such a way as to make possible precise predictions concerning the effects of certain independent variables on recognition-memory performance. Following to some extent the suggestion of Egan (1958), many of the assumptions employed are quite similar to those found by Swets, Tanner, and Birdsall (1961) to be adequate for signal detection in a psychophysical situation.[3]

---

[1] The present paper is adapted from part of a dissertation submitted to the University of Wisconsin. The author wishes to express his sincere gratitude to W. J. Brogden, W. R. Thurlow, M. L. Posner, T. Natsoulas, and D. W. Bessemer for their encouragement and many helpful suggestions.

This research was supported in part by grants from the National Science Foundation and the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

[2] Now with University of California, Davis.

[3] For those readers who are familiar with Swets et al. (1961), the application of their

GENERAL THEORY

Dependent Variables

In taking the recognition test, the subject may be required to choose a certain total number of test-list items (usually equal to the actual number of old items in the test), or he may be free to choose as many or few items as he desires. The present paper is concerned primarily with performance in the latter situation. Since the procedure employed in these experiments does not require the subject to pick some particular number of items, the number of old items that he calls old does not necessarily covary with the number of new items that he calls old. These two scores, expressed as the proportion of each type of item that the subject calls old, are the primary dependent variables of interest.

Fundamental Mechanisms

The present theory posits the existence of a psychological continuum according to which each decision to respond "old" or "new" is made. This dimension may be thought of as representing the degree of familiarity associated with test items and accordingly will be named the familiarity continuum (X). It represents that property of the covert response to the test stimuli which is effective in the determination of the subject's overt report.

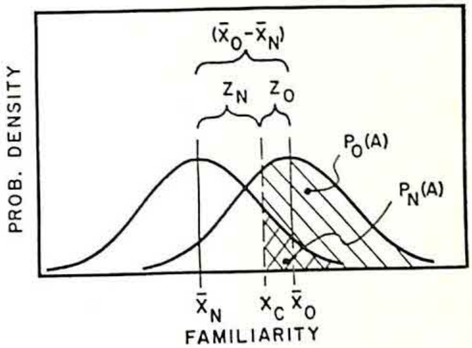

FIG. 1. Basic assumptions: $f_O(X)$, the curve to the right, represents the distribution of events which occurs on those occasions when the subject inspects an old item; $f_N(X)$, the curve to the left, represents the distribution of events which occurs when new items are inspected. (Assuming that the subject sets his cutoff at (X_C) and responds "old item" if and only if (X_i) exceeds (X_C), then ($P_O(A)$) represents the probability of the subject responding old when in fact the item is an old item. ($P_N(A)$) represents the probability of the subject reporting old when in fact the item is a new item.)

When the subject inspects an old item in the test list, it is assumed that a covert event occurs which has some value (X_i) on the familiarity continuum. Furthermore, (X_i) varies from item to item according to a normal distribution, $f_O(X)$. It is further assumed that an event on the same continuum occurs when the subject inspects any of the new items on the test list (as a result of generalization from the old items on transfer of prior learning). These events also form a normal distribution, $f_N(X)$, whose variance is equal to that of $f_O(X)$. The distance on the continuum between the means of the two distributions ($\bar{X}_O - \bar{X}_N$) may be determined in Z units relative to the variance of the item distributions.[4] These assumptions are illustrated in Figure 1.

The quantity ($\bar{X}_O - \bar{X}_N$) is in-

---

theory to recognition memory may be made by considering each test item to be a "trial" with old items being "signal plus noise" trials and new items being "noise only" trials. However, the hypothesized mechanism which determines the position of the cutoff is not equivalent to the optimal strategy rule proposed by Swets et al. Egan in applying Swets et al. to recognition memory was not concerned with the precise determinants of the cutoff nor with the application of the theory to the multiple-choice test situation discussed later.

[4] The measure ($\bar{X}_O - \bar{X}_N$) corresponds to (d') in Swets et al. (1961).

fluenced by learning and retention. These variables determine the position of the old-item distribution on (X). In addition, $(\bar{X}_O - \bar{X}_N)$ may be affected by generalization (which influences the final position of the new-item distribution) and by transfer of prior learning (which affects the initial positions of the new and old distributions). If, however, the same population of old and new items is used in all conditions of an experiment, then variation in $(\bar{X}_O - \bar{X}_N)$ represents variation in learning and retention.

It is further assumed that the subject's overt report regarding item (i) of the test list corresponds to the decision that the event (X_i) associated with item (i) is a member of $(f_O(X))$ or, contrariwise, that it is a member of $(f_N(X))$. We shall assume that this decision is made in terms of a cutoff (X_C) set on (X). That is, the subject will report that item (i) is an old item if, and only if, $(X_i) > (X_C)$ (see Figure 1).

For any given position of the cutoff, the probability that an old item will be reported as old (call this value $P_O(A)$ or the probability of a "*true* positive") is represented by the area under $(f_O(X))$ which lies to the right of (X_C). Similarly the probability that a new item will be called old (call this value $P_N(A)$ or the probability of a "*false* positive") in the area of $(f_N(X))$ to the right of the cutoff (see Figure 1). The proportions of true positives $(P_O(A))$ and of false positives $(P_N(A))$ actually obtained in an experiment are taken to be estimates of $(P_O(A))$ and $(P_N(A))$, respectively.

It follows that if the distance between the distributions is fixed (that is, if learning, retention, generalization, and transfer are constant), then the probability of a false positive may be reduced only by moving the cutoff to the right. Such a change would produce a simultaneous decrease in true positives.

Egan (1958) presented evidence in support of these basic assumptions in a recognition-memory test situation. His procedure involved noting the effects of encouraging or discouraging guessing on the probabilities of true positives $(P_O(A))$ and of false positives $(P_N(A))$. His method for testing these basic assumptions (i.e., plotting "ROC" curves) is equivalent to the following procedure. The proportions of true positives and that of false positives obtained in one of the test conditions are taken as estimates of the probability of true positives, $P_O(A)$, and false positives, $P_N(A)$, respectively. This pair of values is used to estimate the distance $(\bar{X}_O - \bar{X}_N)$ between the old- and new-item distributions (see Figure 1). That is, a table of normal curve will show what the position of the cutoff (in units of Z from the mean of the old-item distribution) *must have been* in order that the area of the curve above the cutoff would be equal to the theoretical $P_O(A)$. Similarly, the value $P_N(A)$ is used to find the distance between the cutoff and the mean of the new-item distribution. These two values are added together to give the estimate of $(\bar{X}_O - \bar{X}_N)$.

When this procedure was repeated for each of the test conditions, it was found that the resulting values for $(\bar{X}_O - \bar{X}_N)$ were essentially constant. Thus, changes in $P_O(A)$ produced by encouraging guessing were accompanied by appropriate changes in $P_N(A)$.

The Cutoff Rule: The Expected-Proportion Model

Beyond these basic assumptions, the present theory attempts to describe exactly *where* the cutoff will be set with respect to the old- and new-item distributions.

It is assumed that the subject is instructed as to the proportion of all the test items that are, in fact, old items. If the cutoff is set too low, the number of items that will be called old will probably be greater than the number of old items in the test. Similarly, if the cutoff is too high, the subject will probably choose fewer items than the actual number of old items. The unique assumption of the present model is that the subject's cutoff will be set at such a place on the familiarity continuum that he will be *expected to choose a proportion of items equal to the proportion of items which are, in fact, old.*[5]

For any given cutoff, the expected proportion of items which a subject will choose is given by the straightforward expression:

$$EP = (P(O))(P_O(A)) + (P(N)(P_N(A)) \qquad [1]$$

where EP equals the proportion of test-list items the subject will be expected to check or, in other words, will be *most likely* to check; $P(O)$ is the proportion of all test items which are, in fact, old; and $P(N)$ is the proportion of all test items which are new.

The basic assumption of this model is that the subject's cutoff will be set such that the expected proportion of items checked (EP) will equal some proportion (K) of all test items which are, in fact, old. That is, it is assumed that:

$$EP = (K)(P(O)) \qquad [2]$$

By substituting Equation 1 into Equation 2, we have the definition of the expected-proportion cutoff rule:

$$(P(O))(P_O(A)) + (P(N))(P_N(A)) = (K)(P(O)) \qquad [3]$$

The only unknown of Equation 3, the value of (K), is assumed to be influenced in an unspecified manner by the costs and rewards of correct and incorrect responses. In experiments in which no deliberate attempt is made to vary these values, (K) is assumed to be constant. The fundamental assumptions and the expected-proportion rule may be tested separately or may be combined and tested simultaneously. This latter procedure will be illustrated shortly.

EVIDENCE FOR THE EXPECTED-PROPORTION MODEL

The Davis, Sutherland, and Judd Experiment

Davis, Sutherland, and Judd (1961) examined recognition-memory performance using both consonant-vowel nonsense syllables and two-digit numbers as stimuli.[6] The data for the nonsense-syllable condition will be examined first.

Each nonsense syllable consisted of 1 of 18 consonants (omitting q and z) followed by 1 of 5 vowels. A random list of 15 items was presented to the subject as the inspection list, each item being presented for 1.5 seconds. Immediately following the presentation of the inspection list, the subject was given a sheet of paper containing the 15 inspection (old) items together with 15, 45, or 75 new items (the test list). Thus, the proportion of old items, $P(O)$, in the test list was either .500, .250, or .167. There were 24 subjects in each condition. The instructions, which were the same in all conditions, required the subject to check as many old items as he could but not to guess wildly. The fundamental dependent variables, for our purposes, are the proportion of true positives ($p_O(A)$) and the proportion of false positives ($p_N(A)$) obtained in each condition. These values appear in Figure 2.

---

[5] The word "expected" is used in the mathematical sense. The expected proportion of outcomes of a certain sort (in this case, of an item being called old) is the proportion which is *most likely* to occur over a large sample of occasions (in this case, a large number of test items). The expected proportion may be found simply by multiplying the probability of the outcome (the response) by the proportion of occasions (items).

Although the subject is not *required* to choose some fixed proportion of the items, we assume that he behaves in such a way as to make it *most likely* that he will choose a certain proportion of items.

[6] They also tested recall performance with these same stimuli, but only the recognition data is of interest here.

The test of applicability of the expected-proportion model to these data is relatively direct. Since the conditions of learning and recognition are constant (except for the values of $P(O)$ and $P(N)$), the difference between the means ($\bar{X}_O - \bar{X}_N$) and the value of (K) are presumably constant. The values of $p_O(A)$ and $p_N(A)$ for any one group can be used to estimate both of these parameters.

The data obtained under the $P(O) = .50$ condition were arbitrarily chosen for this purpose. In this group $P_O(A)$ was found to be (.694). A table of the normal curve shows that the probability of a true positive would be (.694) if the cutoff were located at (.51) Z units below the mean of the new-item distribution. Similarly, the probability of a false positive would be equal to the obtained value of (.186) if the cutoff were located at (.89) Z units above the mean of the new-item distribution. Adding these distances (see Figure 1) we estimate that the distance between the means ($\bar{X}_O - \bar{X}_N$) is equal to (1.40) Z units.

Now in order to estimate (K), these same values of $P_O(A)$ and $P_N(A)$ are placed in Equation 3 along with values of $P(O) = P(N) = .50$. Solving the equation, (K) equals .880.

Since both of these parameters are assumed to be constant, we assume that the subjects in each of the other conditions are faced with two normal distributions at a distance of (1.40 Z). Furthermore, we assume that their cutoff will be set such that they will be expected to check a proportion of items equal to (.880) times the value of ($P(O)$) for their conditions. Thus, for example, when $P(O) = .25$, they will attempt to check (.880), (.25), or (.220) of all the test-list items.

On the basis of these assumptions we may predict precisely the values of $P_O(A)$ and $P_N(A)$ which should be obtained in each of the other conditions. This procedure involves a trial and error search for the place where the cutoff *should* be set in order that Equation 3 will be satisfied.[7] The

---

[7] A cutoff point in terms of the distance (Z_O) between the cutoff and the mean of the old distribution is *tentatively* selected, and the resulting tentative value of $P_O(A)$ is found from a table of the normal curve. Since the distance between the means is assumed to be (1.40 Z), there will be a corresponding tentative value of Z_N and of $P_N(A)$. These tentative values of $P_O(A)$ and $P_N(A)$ are used in Equation 3 along with the values for $P(O)$ and $P(N)$ for the condition. This procedure is repeated until a cutoff position is found such that the expected proportion equals (.880) ($P(O)$). This procedure may be illustrated for the $P(O) = .250$ condition as follows:

Tentative Z_O	Tentative $P_O(A)$	Corresponding Z_N	Tentative $P_N(A)$	Expected proportion
−.20 Z	.579	1.20 Z	.115	.231 Reject
−.15 Z	.560	1.25 Z	.106	.220 *Accept*
−.10 Z	.540	1.30 Z	.097	.208 Reject

Since .220 equals (.880) (.250), we accept the prediction that the cutoff will be located .15 Z units below the mean of the old-item distribution and that, therefore, $P_O(A)$ will equal .560, and $P_N(A)$ will equal .106.

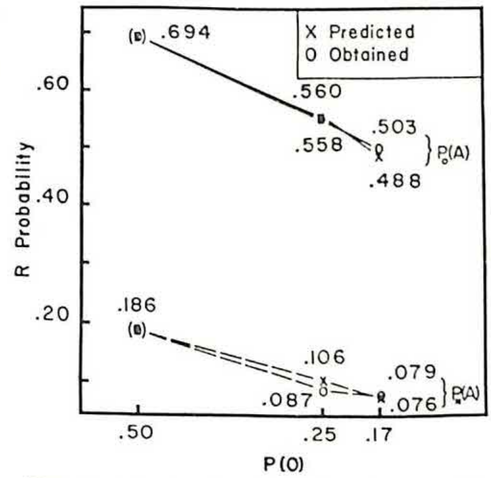

FIG. 2. Obtained and predicted recognition data when nonsense syllables are used as stimuli; $K = .880$; $(\bar{X}_O - \bar{X}_N) = 1.40Z$ (Davis, Sutherland, & Judd, 1961). (In this and the following figures, the data in parenthesis were used in making the predictions and, therefore, must coincide with the predictions.)

cutoff which yields values of $P_O(A)$ and $P_N(A)$ that satisfy this equation is taken as the predicted place of the cutoff, and these values of $P_O(A)$ and $P_N(A)$ are taken as the predictions of $p_O(A)$ and $p_N(A)$, respectively. The resulting predictions are compared to the obtained data in Figure 2.

In another experiment reported in the same paper Davis et al. (1961) used numbers as stimuli. These items consisted of all possible permutations of 10 digits taken 2 at a time. In all other aspects the experiment was identical to the one given above, and consequently the analysis is the same. Once again, the data obtained when $P(O) = .50$ is used to estimate (K) and $(\bar{X}_O - \bar{X}_N)$, and these values are used to predict the data under $P(O) = .250$ and $P(O) = .167$.

The Davis et al. experiments, while providing excellent support for the expected-proportion model, leave certain points unclarified.[8] The entire test list was exposed, and the subject was allowed to respond to the items in any order he wished. Under these conditions the subjects can produce data which are consistent with the expected-proportion model without, technically, using a cutoff at all (although the other, more fundamental assumptions would need to be valid). That is, the subject, faced with the two normal distributions, could start checking off the items with highest familiarity and work down to items with lower familiarity and simply stop when he had checked a number of items approximately equal to the number of old items on the test list. Such a strategy, on the other hand, would be impossible if the subject were allowed to view only a single test item at a time and was required to make his decision on that item before moving on to the next item. A demonstration that the expected-proportion model is adequate

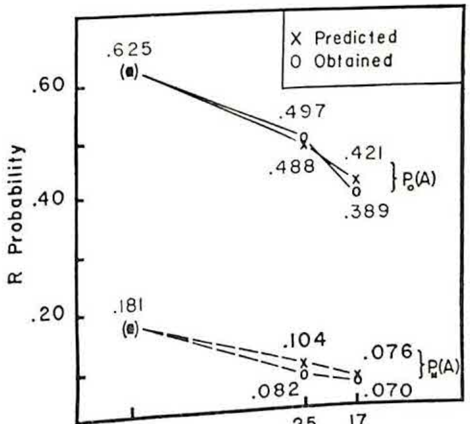

FIG. 3. The predicted and obtained recognition-memory data when two-digit numbers are used as stimuli; $K = .806$; $(\bar{X}_O - \bar{X}_N) = 1.23\ Z$.

---

[8] The goodness of fit of predicted to obtained data is ascertained by inspection throughout this paper except in Experiments I and II described later. The test used in the latter cases requires data for individual subjects. These data are not available for the other experiments that are discussed.

under such a serial-test procedure would provide less ambiguous support for the notion of a cutoff. Two studies by the present author were designed to shed some light on this question. The method and results of these studies are described in detail in Parks (1964).

Experiment I

A total of 10 groups of 8 subjects each was presented with a 16-item list of consonant-vowel-consonant trigrams of low association value (Archer, 1960). Each subject viewed each item of the list only once, the exposure of each item being 1.67 seconds with a 1.33-second break between items. All subjects received the same inspection list, but eight different orders of presentation were used within each group. Immediately following the presentation of the last inspection item, the subject was presented with a test list composed of a random mixture of the 16 inspection items and 4, 8, 16, 32, or 64 new consonant-vowel-consonant trigrams. Thus, the proportion of test items that were old items was either $P(O) = .80$, .67, .50, .33, or .20. In each case the subject was carefully instructed as to the number of old and number of new items which would appear in his test list. Care was taken to balance the effects of the particular new items used across the levels of $P(O)$.

In addition to differences in $P(O)$, the tests differed in response procedures. Half of the groups were allowed to uncover the entire list at the beginning of the test and to respond in any order they desired ("free-responding" condition). The other five groups had a series of sliding masks over their test lists such that they could view only a single item at a time and were forbidden to return to an item already examined ("serial-responding" condition).

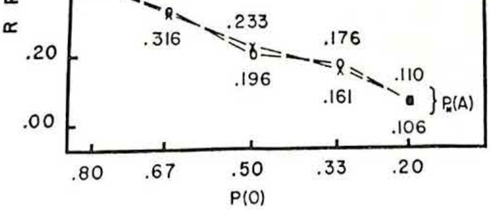

FIG. 4. Obtained and predicted data for the free-responding condition of Experiment I. (Predictions are made from the expected-proportion model and the $P(O) = .80$ data; $K = .977$; $(\bar{X}_O - \bar{X}_N) = 1.39\ Z$.)

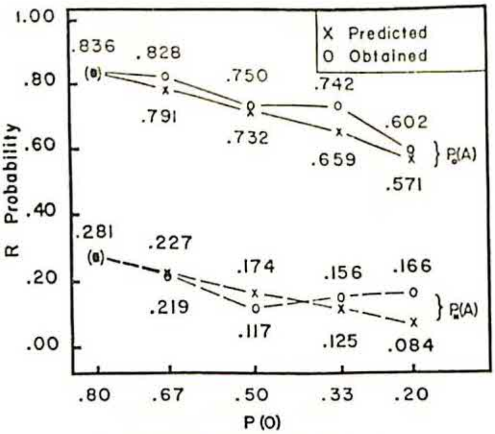

FIG. 5. Obtained and predicted data for the serial-responding condition of Experiment I. (Predictions are made for the expected-proportion model and the $P(O) = .80$ data; $K = .906$; $(\bar{X}_O - \bar{X}_N) = 1.56\ Z$.)

In both cases the subjects were instructed to underline any item which they thought had been on the original list. They were permitted to guess but encouraged not to guess wildly. The actual lists used in the two responding conditions were identical except for the masks placed over the lists in the serial condition. Finally, in both conditions, the subjects were free to take as much time as they desired but were encouraged to "keep moving along."

The treatment of the data is identical to that employed above with the data of Davis et al. The $P_O(A)$ and $P_N(A)$ scores obtained in $P(O) = .80$ under the free-responding condition are used to estimate $(\bar{X}_O - \bar{X}_N)$ and (K) and, from these values, to predict the remaining free-responding data. The mean data and predictions appear in Figure 4. Similarly, the serial-responding $P(O) = .80$ data are used to predict the performance of serial-responding groups (see Figure 5).

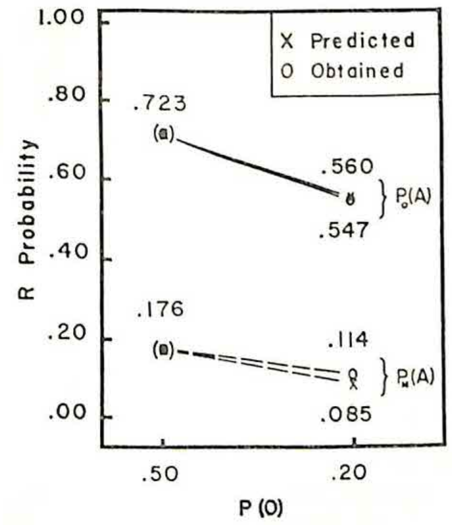

FIG. 6. Obtained and predicted data of Experiment II in which the $P(O)$ of the first few items was equal to the overall $P(O)$ for the condition. (Predictions are made from the expected-proportion model and the $P(O) = .50$ data; $K = .899$; $(\bar{X}_O - \bar{X}_N) = 1.52$.)

The goodness of fit of each of the four predicted curves was evaluated using a test suggested by Grant (1962). Essentially this procedure tests the correlation of predicted and obtained means against the variance within conditions. The null hypothesis is that the correlation is no greater than that which would be expected on the basis of chance. The results are as follows: free responding, $P_O(A)$: $F = 24.068$; free responding, $P_N(A)$: $F = 17.716$; serial responding, $P_O(A)$: $F = 19.826$; and serial responding, $P_N(A)$: $F = 5.030$. In each case $df = 1/35$. All of these are significant at the .05 level of confidence. In fact only the serial-responding $P_N(A)$ measure fails to achieve significance at .01.

Inspection reveals that under the serial-responding condition the obtained value of $p_N(A)$ is lower than predicted at $P(O) = .50$ and higher than predicted at $P(O) = .20$. Reexamination of the test lists employed in these conditions suggested that this effect might have been due to the proportion of old items among the first few items of the test list. It was found that the mean proportion of old items in the first 10 items for the $P(O) = .50$ lists was .60. The mean proportion among the initial 10 items for the $P(O) = .20$ lists was .10. These differences between overall $P(O)$ and $P(O)$ within the initial test items might well have made the subjects more conservative in the former case and more liberal in the latter. This variable would not systematically affect the free-responding subjects since they viewed the items in any order desired. A second experiment supports this hypothesis.

Experiment II

Two groups of eight subjects each were tested under identical conditions to those of the serial-responding $P(O) = .50$ and $P(O) = .20$ groups of Experiment I. In this case, however, the test lists employed were randomly constructed *but* with the restriction that the $P(O) = .50$ lists each have 5 old items among the first 10 test items and lists for $P(O) = .20$ each have 2 old items among the first 10 items.

The results of this study appear in Figure 6 along with the theoretical curve generated from the $P(O) = .50$ data. In this case the F value for correspondence of obtained and predicted means is equal to 12.598 for $P_O(A)$ and 5.000 for $P_N(A)$ ($df = 1/14$, $p < .01$, and $p < .05$), respectively.

In these experiments, as in those of Davis et al., when $P(O)$ is varied, the subjects' performances varied as though they were using the expected-proportion cutoff rule. Furthermore, the fit of the model to the serial-responding data indicates that the cutoff hypothesis may indeed be valid and lends support, in the interests of parsimony, to the cutoff type of description of free-responding performance.

Finally, it is interesting to note that the theoretical curves which fit these data are based, in part, on the assumption that $(\bar{X}_O - \bar{X}_N)$ is constant. According to the theory, this fact may be interpreted as indication that retention was constant across the various levels of $P(O)$ even though the subjects in the lower levels of $P(O)$ took more time to complete the test. In the $P(O) = .80$ condition, the subjects required an average of 112 seconds and 121 seconds to complete the test in the free-responding and serial-responding groups, respectively. When $P(O) = .20$, a mean of 394 and 273 seconds was required. A 2 × 5 analysis of variance on the time measures revealed a significant effect of $P(O)$ level ($F = 12.901$, $df = 4/70$, $p < .01$).

From a methodological standpoint, this finding illustrates the value and even necessity of precise theory since it could not have been ascertained by simple examination of the data. Perhaps, in fact, the main value of the present theory lies in its being able to provide a means of evaluating retention which allows for, rather than being confounded with, the particular conditions under which retention is measured.

Before proceeding, one possible source of confusion regarding these tests of the expected-proportion model must be discussed. If the subjects tend to choose a certain fixed number of items, the resulting pair of $p_O(A)$ and $p_N(A)$ values will necessarily satisfy Equation 3. But there are a variety of such pairs that may occur. The accuracy of the predictions given above demonstrates that the particular pair that occurs in each condition is that *unique* pair of areas that *both* (*a*) will satisfy Equation 3 and (*b*) *may occur* when a cutoff is moved along two normal curves of equal variance at a fixed distance.

Other Applications of the Expected-Proportion Model

The expected-proportion model may also be tested by holding $P(O)$ and $P(N)$ constant while some other variable of learning or retention is manipulated. In this case, however, as we shall see, only the ideas expressed in Equation 3 are tested (i.e., no test is made of the fundamental assumption of normality).

One such experiment has been reported by Strong (1912). The number of stimuli in the presentation list was the independent variable. Either 5, 10, 25, 50, 100, or 150 magazine advertisements were presented one at a time at a rate of one per second. Immediately following the presentation of the inspection list, the subject was given a stack of stimulus cards containing a random mixture of the inspection items (i.e., "old" items) together with an equal number of new advertisements. Thus, the test list contained either 10, 20, 50, 100, 200, or 300 items. The subject was required to sort the test items into four stacks according to his degree of conviction that each item was an old item. For our purposes, however, we will lump together the three classes that included items which the subject "had at least a faint idea he had seen before . . . i.e., those of which he was about 25% sure [p. 449]." The proportion of the old items in the test list which the subject placed into any one of these three classes shall be considered to be the proportion of true positives ($p_O(A)$). The proportion of all new items in the test list which the subject placed into any of these three classes is taken as the proportion of false positives ($P_N(A)$).

The values of the parameters ob-

TABLE 1

PROPORTION OF TRUE POSITIVES AND FALSE POSITIVES OBTAINED BY STRONG (1912) AS A FUNCTION OF THE NUMBER OF ADVERTISEMENTS IN THE INSPECTION LIST

Condition	No. *Ss*	No. old list items	No. old list items	Obtained $P_O(A)$	Obtained $P_N(A)$
Rec 10	80	5	5	.891	.036
Rec 20	40	10	10	.885	.063
Rec 50	40	25	25	.823	.074
Rec 100	40	50	50	.726	.114
Rec 200	40	100	100	.678	.135
Rec 300	40	150	150	.530	.129

Note.—Test list consisted of the inspection-list items together with an equal number of new items.

tained in each condition are given in Table 1 along with the number of subjects for each condition.

Ordinarily, the expected-proportion model might be tested against these data by simply placing the estimates of $P(O)$ and $P(N)$ obtained in each group along with values $P(O) = P(N) = .50$ into Equation 3 and solving for (K). The value (K) should, of course, be constant. However Equation 3 is based on the assumption that the subject has been informed of the values, $P(O)$ and $P(N)$. According to his report, Strong's subjects were apparently not so instructed. Thus, the model must be modified as follows:

$$(K)(P(O))' = (P(O))'(P_O(A)) + (P(N))'(P_N(A)) \quad [4]$$

where these prime marks indicate values assumed by the subject rather than the actual values of the parameters. That is, the prime values are fitting parameters.

By one interpretation this equation simply states that the subject will set his cutoff such that the expected proportion of items he *believes* he will check (based on his estimates of $P(O)$ and $P(N)$) will be equal to some constant fraction of what he *believes* to be the proportion of test-list items which are old items.

In order to test this model against Strong's data, Equation 4 must be rewritten as follows:

$$(K) = (P_O(A)) + \frac{P(N)'}{P(O)'}(P_N(A)) \quad [5]$$

Now if it is assumed that $P(O)'$ and $P(N)'$ are constant across groups, then $P(N)'/P(O)'$ will be constant and may be treated as a single fitting parameter. As usual, (K) is also assumed to be constant. Since $P_O(A)$ and $P_N(A)$ are known for each group, an equation in two unknowns may be written from the data of each group. Solving two such equations (from the 10 and 20 conditions) simultaneously, it is found that $P(N)'/P(O)' = (2.22)$ and $(K) = (.971)$.

It is now possible to predict the value of $P_N(A)$ for any group by placing these values together with the estimate of $P_O(A)$ for that group in Equation 5. These predictions are compared to the obtained values in Figure 7. Except for data obtained with the 300-item list, the fit appears to be satisfactory.

The expected-proportion model may also be applied to data gathered using what has been called a multiple-choice or forced-choice testing procedure. In this case, the test list is divided into subsets consisting of one old item and

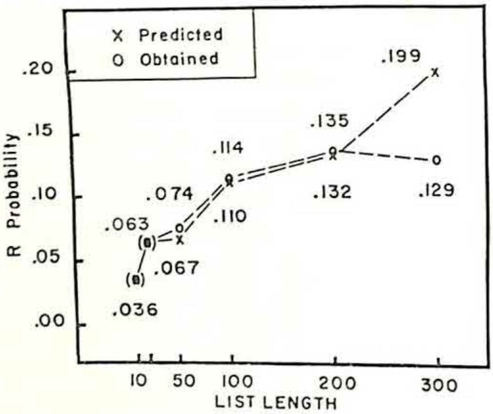

FIG. 7. Obtained and predicted values of $P_N(A)$ using the modified expected-proportion model and obtained $P_O(A)$ in each condition. (This experiment differed from Davis et al. in that (*a*) $P(O)$ and $P(N)$ were not varied; and (*b*) the subject did not know the value of $P(O)$—Strong, 1912.)

(n) new items. The instructions employed along with this procedure usually require the subject to choose no more nor less than one item out of each subset. This requirement ordinarily would place the experiment beyond the scope of the present theory. That is, if the subject is required to examine a subset *and* to pick one item *before proceeding* to the next set, he cannot use a cutoff at all since it may be that no item exceeds his cutoff, and yet he must respond.[9] On the other hand, the cutoff rule may be employed in a multiple-choice situation if the subject is free to move back and forth among the sublists. In this case, if no item exceeds the cutoff in one subset, the subject may go on to another subset and return later.

In order to apply the expected-proportion model to this sort of data, all of the assumptions of the expected-proportion model are made. That is, it is assumed that the subject is faced with two normal distributions at fixed distance on a familiarity continuum and that the cutoff is set according to the expected-proportion model. Furthermore, the additional assumptions are made that (*a*) in each subset the subject will choose the first item that he examines which exceeds the cutoff, and (*b*) that (K) = 1.[10] According to Assumption *a* the probability of a true positive (estimated by $p_O(A)$) will be given by:

$$p_O(A) \simeq \frac{\text{expected number of old items that exceed cutoff}}{\text{expected total number of items that exceed cutoff}} \qquad [6]$$

The expected number of old items that exceed the cutoff is equal simply to the probability that the single old item in each subset exceeds the cutoff—that is, it is equal to $P_O(A)$. The expected number of new items that exceed the cutoff is given by the probability that any particular new item exceeds the cutoff—that is, by $P_N(A)$—multiplied by the number of new items (n) in the subset. Therefore, Equation 6 may be rewritten as:

$$p_O(A) \simeq \frac{(1)\ (P_O(A))}{(1)\ (P_O(A)) + (n)(P_N(A))} \qquad [7]$$

---

[9] In this case the subject would probably simply pick the item with the greatest familiarity response. In support of this hypothesis, it can be shown that the raw (uncorrected) data reported by Flores (1958) may be fit by the equation $P_O(A) = P(X_O > X_N)^n$ where (n) is the number of new items in each subset, and $P(X_O > X_N)$ is the probability that one old item will exceed any one new item. This latter quantity is treated as a fitting parameter and is assumed to be constant if the conditions of learning, retention, and test (except n) are constant.

[10] For our purposes this assumption is equivalent to the assumption that the subject rejects some members of each subset (those that do not exceed the cutoff) and makes a random guess among the remaining items. Murdock (1963) has suggested and defended

Now if $(K) = 1$, Equation 3 given earlier may be rewritten as:

$$P_N(A) = \frac{1 - P_O(A)}{n}$$

Substituting this expression into Equation 7 and simplifying:

$$p_O(A) \simeq P_O(A) \qquad [8]$$

In other words, it follows from the assumptions, that the obtained proportion of true positives will be, as usual, an estimate of the probability that an old item will exceed the cutoff. Similarly, it can be shown that $p_N(A)$ estimates $P_N(A)$.[11] Thus, the model may be tested in exactly the same manner in which it was tested against the data of Davis et al. In this case, of course, (K) is not a fitting parameter but is assumed to equal one.

It should be noted that in testing the theory against multiple-choice data, only the fundamental assumptions of normality and a movable cutoff are being tested. That is, it is known a priori that the data will satisfy Equation 3 since the subject is forced by the instructions to choose a constant number of items (i.e., one per subset). On the other hand, although a variety of pairs of $p_O(A)$ and $p_N(A)$ values will satisfy this requirement, there is a unique pair which will also satisfy the requirement of representing corresponding areas under two normal distributions at a fixed distance.

The Postman Experiment

An experiment reported by Postman (1950) used this paradigm. He presented a list of 36 nonsense syl-

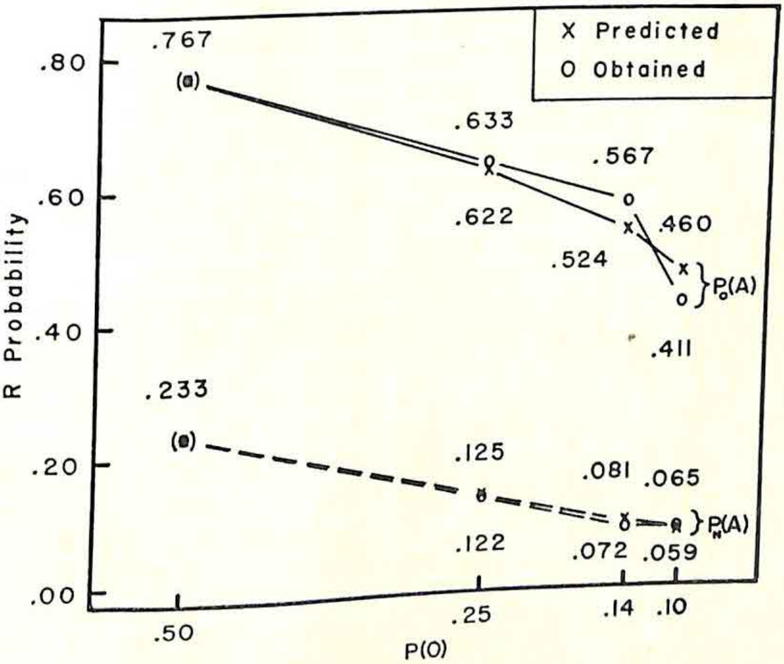

FIG. 8. Obtained and predicted data using a modified expected-proportion model; $(\bar{X}_O - \bar{X}_N) = 1.46$ (Postman, 1950).

---

a hypothesis which is very similar to this latter assumption.

[11] The score $p_N(A)$ is the proportion of new items that are called old. Therefore, it is equal to the proportion of subsets on which a new item is chosen divided by the number of new items in each subset.

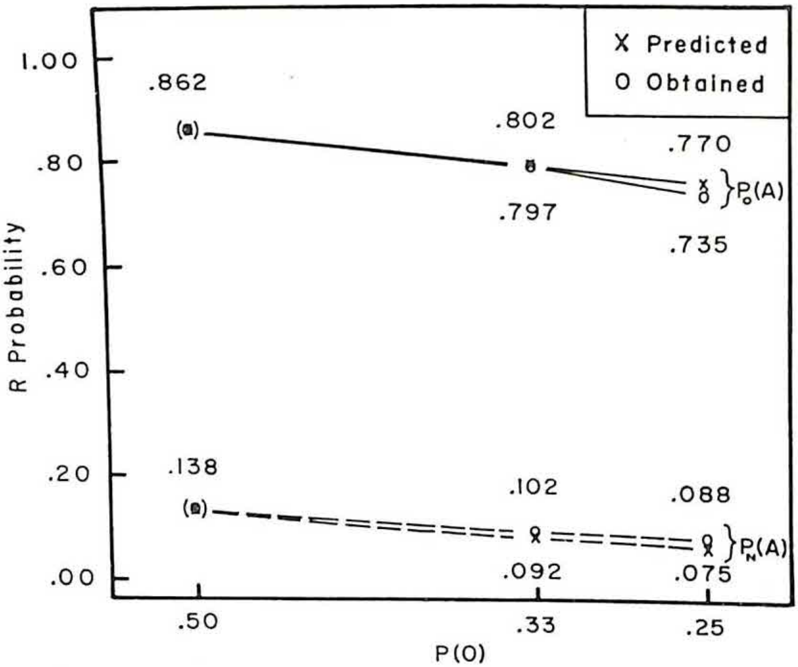

FIG. 9. Obtained and predicted data using a modified expected-proportion model; $(\bar{X}_O - \bar{X}_N) = 2.18$ (Murdock, 1963).

lables, the list being presented five times in five different orders. Following these learning presentations, a written multiple-choice recognition-memory test was administered. This test consisted of a series of 36 subsets, each subset containing one old item and either one, three, six, or nine new items. Thus, there are four recall conditions; $P(O) = .50$, .25, .14, and .10, respectively. These 36 multiple-choice subsets were presented on a single page, and the subject was asked to indicate which member of each list was an old item.

The data obtained under the $P(O) = .50$ condition are used to estimate $(\bar{X}_O - \bar{X}_N)$. This value along with $(K = 1)$ and level of $P(O)$ are used to predict the rest of the data in the manner described for the Davis et al. experiment. The obtained and predicted values of $p_O(A)$ are given in Figure 8. The fit appears to be satisfactory.

The Murdock Experiment

Murdock (1963, p. 18) reports a very similar experiment. Using general knowledge statements as stimuli, a list of 200 multiple-choice subsets was administered to 97 subjects. Of these 200 subsets, 67 contained one correct statement ("old" item) and one incorrect statement ("new" item), 67 contained one old and two new items, and 66 contained one old and three new items. Thus, $P(O) = .50$, .33, or .25. The instructions required the subject to select one item of each subset. The obtained values of $p_O(A)$ and $p_N(A)$ after Murdock's correction for guessing is removed are given in Figure 9 along with the predicted values. The predictions are made from the $P(O) = .50$ data (where (K) is assumed to equal 1). The fit appears to be satisfactory.

The studies reviewed in this section are especially important in distinguishing the movable-cutoff model underlying the present theory from the sophisticated fixed-threshold model proposed by Luce (1963) for psychophysical detection. Luce's model, when suitably translated for application to the recognition-memory situation, may be shown to be consistent with the data obtained by Davis et al. (1961), Parks (1964), and Strong (1912).[12] However, this fixed-threshold model is apparently incapable of explaining the precise effect of $P(O)$ on performance under the multiple-choice procedure of Postman (1950) and Murdock (1963). Although the data of these latter experiments are consistent with the very broad predictions yielded by Luce's "K alternative" model, there is no ready explanation of the fact that the obtained data lie in each case on a single ROC function.

CONCLUDING COMMENTS

Of the independent variables which occur in the studies reviewed, only those concerned with test procedure are specifically handled by the theory. The other variables which have been mentioned (length of list, type of stimuli, and procedure) are indifferent with respect to the theory. That is, they may or may not be effective variables of $(\bar{X}_O - \bar{X}_N)$ and (K). In this sense, the present theory is concerned with recognition-memory performance rather than learning and retention per se. Nevertheless, in describing the end product of learning and retention, the theory has implications for any description of that learning upon which recognition responses are based. On a broader scale, the great similarity of the present theoretical assumptions to those which have been found to be adequate in psychophysical situations suggests that the essential features of this theory might, in time, be shown to be only a special application of a very broad theory of decision behavior.

REFERENCES

ARCHER, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **17** (10, Whole No. 497).

DAVIS, R., SUTHERLAND, N. S., & JUDD, B. R. Information content in recognition and recall. *Journal of Experimental Psychology*, 1961, **61**, 422–429.

EGAN, J. P. Recognition memory and the operating characteristic. Technical Report No. AFCRC TN-58-51, AD-152650, 1958, Hearing and Communication Laboratory, Indiana University.

---

[12] Rather than assuming that (t) and (u) are determined by an asymptotic learning process as described by Luce, it is assumed that these biasing factors result from the subject's attempt to match the expected proportion of old responses to $P(O)$. Thus, for example, (t) is set such that:

$$(t)\,(q(O))\,(P(O)) + (t)\,(q(N))\,(P(N)) = K\,(P(O)).$$

That is:

$$t = \frac{K}{q(O) + \dfrac{P(N)\,q(N)}{P(O)}}$$

where K is determined, in an unspecified manner, by the cost and values involved. It follows that:

$$p_O(A) + \frac{P(N)}{P(O)}\,(p_N(A)) = K$$

If, as in Strong's procedure, the subject employs estimates $P(O)'$ and $P(N)'$, then:

$$p_O(A) + \frac{P(N)'}{P(O)'}\,(p_N(A)) = K$$

These predictions are, as far as they go, identical to those made from the expected-proportion, movable-cutoff theory.

The two theories differ in that the sophisticated fixed-threshold theory predicts that, if the conditions of learning and retention are constant, the obtained values $p_O(A)$ and $p_N(A)$ will lie on an ROC curve composed of two intersecting straight-line segments. In practice such a function has been indistinguishable from the Gaussian ROC function of the movable-cutoff theory.

FLORES, C. Etude sur les relations entre le rappel et la reconnaissance. *L'Annee Psychologie*, 1958, **58**, 365–375.

GRANT, D. A. Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review*, 1962, **69**, 54–61.

LUCE, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, **70**, 61–79.

MURDOCK, B. B. An analysis of the recognition process. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning*. New York: McGraw-Hill, 1963. Pp. 10–21.

PARKS, T. E. *An experimental and theoretical analysis of recognition memory performance.* (Doctoral dissertation, University of Wisconsin) Ann Arbor, Mich.: University Microfilms, 1964, No. 64–12741.

POSTMAN, L. Choice behavior and the process of recognition. *American Journal of Psychology*, 1950, **63**, 576–583.

STRONG, E. K. The effect of length of series upon recognition memory. *Psychological Review*, 1912, **19**, 447–462.

SWETS, J. A., TANNER, W. P., & BIRDSALL, T. G. Decision processes in perception. *Psychological Review*, 1961, **68**, 301–340.

(Received September 16, 1964)

Psychological Review
1966, Vol. 73, No. 1, 59–86

HEDONIC ORGANIZATION AND REGULATION OF BEHAVIOR[1]

PAUL THOMAS YOUNG[2]

University of Illinois

The paper is a critical review supplementing previous articles upon the role of hedonic processes in motivation and learning. Special emphasis is placed upon the organizing and regulatory functions of affective processes. Recent experiments from the author's laboratory are reviewed and their theoretical significance considered. These experiments deal with hedonic processes as related to gustatory stimulation, food preference and intake, regulation of ingestive behavior, development of incentive values, and related matters. The main topics considered are: basic concepts and definitions; psychophysical studies of hedonic processes; studies of thresholds, preference, and the regulation of food intake; physiology of affective processes; the role of affective processes in motivation and learning; general conclusions.

In a previous paper (Young, 1959) I reviewed experiments based upon the hypothesis that affective processes have an objective existence within the nervous system. The present paper reports further research within the framework of experimental hedonism. It brings up to date the account of findings and hypotheses. Earlier theoretical papers and bulletins are available (Young, 1941, 1948, 1949a, 1952, 1955, 1957).

BASIC CONCEPTS AND DEFINITIONS[3]

Operational Definitions

Food acceptance is a complex process regulated by at least four groups of determinants: (*a*) *Organic conditions,* including metabolic needs produced by specific deprivations, and special organic states such as pregnancy, lactation, experimental states induced by surgical operations, implanted electrodes, drugs, etc.; (*b*) *peripheral stimulations* of the head receptors (taste, smell, touch, vision, audition) including environmental stimulations from the surroundings of a food. Perhaps under this heading one can include ecological factors such as availability of a foodstuff in the environment, physical condition of food, and human social determinants—economic conditions, religious taboos, esthetic factors, verbal suggestions, etc.; (*c*) *previous experience,* including the dietary history of an organism, the feeding habits formed, attitudes, motives, expectancies, and similar dispositions; (*d*) *bodily constitution,* or the biochemical make-up of organisms which differs from individual to individual and from species to species. It is known, for example, that cats do not taste and discriminate sugar solutions and that birds are weak in the so-called "sweet tooth."

---

[1] This paper is an outgrowth of research supported by National Science Foundation Grants G-14538 and GB-215. I am indebted to W. E. Kappauf for reading this paper and making many helpful suggestions.

[2] Retirement address: 336 Notre Dame Road, Claremont, California 91712.

[3] Throughout this paper I will use, for simplicity and convenience, the symbols and specifications employed in recent experiments as follows: CHO = sucrose; NaCl = sodium chloride; Q = quinine hydrochloride; H = tartaric acid; Sac = sodium saccharin; H_2O = distilled or deionized water. Unless otherwise indicated, concentrations will be expressed as grams of solute dissolved in 100 cubic centimeters of solution. Thus 4 CHO designates a simple sucrose solution containing 4 grams of sucrose per 100 cubic centimeters of solution; .01 Q + 1.59 CHO designates a compound solution containing .01 gram of quinine hydrochloride plus 1.59 grams of sucrose per 100 cubic centimeters of solution (see Pfaffmann, Young, Dethier, Richter, & Stellar, 1954).

Referring to these groups of determinants we define *drive, appetite, specific hunger, general hunger,* etc., as organic motivations regulating behavior when the other groups of determinants are held constant. We define *palatability* as the hedonic characteristic of a food dependent upon stimulation of the head receptors by a foodstuff and/or its surroundings when other determinants are held constant. We define *food habit* as food acceptance determined by previous experience rather than by the other conditions. We speak of *constitutional determinants* of food acceptance and study individual and species differences when other factors are controlled. Basic concepts can thus be defined operationally.

The terms *acceptance* and *acceptability* are neutral and carry no subjective stigma. These terms refer to observed ingestive behavior and are free from bias as to conditions that determine food intake.

The Concepts of Choice and Preference

A distinction should be drawn between choice and preference. *Choice* implies that there are two or more possible courses of action for which an organism is primed. *Preference* is a form of choice which carries the additional implication that one course of action is better than another. Preference is an evaluative concept implying that Response A is more acceptable, more desirable, more esteemed, more valued, better liked, more pleasant than Response B. A choice may be made without evaluation since many factors determine choice.

A serious question can be raised as to the extent to which affective processes are actually involved in choice. According to Lloyd Morgan's canon, we should not postulate affective determinants of choice if choice can be explained completely in terms of sensory stimulations, reflexes, and habits.

The necessity of distinguishing between habit and palatability as separate determinants of choice came to light in an early experiment by Young and Chaplin (1945). They observed two incompatible choices in quick succession with the same rats. On the standard preference tester the rats showed a persisting preference of sugar to casein. This was the preference that originally developed under need-free conditions. The preference persisted, however, during a period of steadily increasing protein starvation. When the protein-starved rats were tested on a different kind of apparatus, they promptly developed a new and incompatible preference (casein > sugar). Now the concept of two simultaneous but incompatible preferences does not make sense. So control experiments were made (Young, 1945). In the light of the control experiments I interpreted the persisting sugar preference in the original experiment as a preferential habit that did not correctly reveal the changed palatability relations, and I concluded that preferences tend to develop in agreement with metabolic need states, but preferential habits persist as independent determinants of choice even though the choice is unwise—contrary to metabolic needs.

It is clear from such evidence that choice has various determinants. One of these is the *hedonic outcome* of a response. I have postulated on sound grounds that hedonic processes regulate contemporary behavior according to the principle of maximizing the positive and minimizing the negative, and that hedonic processes organize neural dispositions which persist and regulate behavior as independent determinants of choice. The hedonic outcome of a choice regulates the development of preferential discriminations. For example: A naive rat approaches the choice point on a T maze; he can turn right or left. On Trial 1 he may turn either way depending on a variety of predisposing factors, but after he has experienced the taste of sugar in the left goal box and the taste of casein (or perhaps nothing) in the right goal box, the probability that he will turn left on successive runs increases approaching the asymptote of 100% left turns. We conclude that the rat prefers sugar to casein. The preference is revealed by the increasing probability of a left turn which we have called the preferential trend.

Therefore, it is clear that preference has at least two meanings. First, preference is a *functional* variable—an immediate evaluation based in part upon affective processes and revealed in a single trial. Second, preference is a *dispositional* variable—an acquired habit of discrimination that develops with repeated choices. These two meanings are closely interrelated: Affective processes organize persisting dispositions which in turn regulate choice. A specific choice depends upon existing habits as well as concomitant hedonic processes and possibly other factors.

Experimental Methods with and without Choice

The single-stimulus method was proposed by Beebe-Center, Black, Hoffman, and Wade (1948) as a technique for studying "preference." The method has many useful applications in studies of food acceptance, but it is *not* a true preference method because it excludes the possibility of choice. The single-stimulus method and the running methods clearly reveal differences in performance.

The single-stimulus method was used by Weiner and Stellar (1951) in studying acceptance rates of NaCl solutions. They found that naïve rats gave the normal "preference-aversion" (acceptance-rejection) curves with a maximal rate of acceptance at .8 NaCl and an "aversion point" (indifferent concentration) at about 1.5 NaCl. The subjects did not regulate intake to obtain a constant quantity of NaCl, as required by the homeostatic doctrine, but the quantity ingested varied with NaCl concentration. These findings with the single-stimulus method agree with results obtained by methods that involve choice.

Young and Greene (1953) found that with a single-stimulus method rats ingested more 9 CHO than 36 CHO during a 1-hour drinking test, but when offered a choice they ingested more 36 CHO. Also on the brief-exposure preference tester the subjects revealed a clear and consistent preference of 36 CHO to 9 CHO. Hence preferences cannot always be predicted correctly on the basis of results obtained with the single-stimulus method, and a distinction must be drawn between preference and intake.

Carpenter (1958) offered rats a choice of CHO and NaCl solutions with a two-bottle and an eight-bottle technique. He found that the pattern of intake differed with the method of presentation. When the rats had choice among eight CHO solutions, the data showed that absolute and relative intake varied directly with concentration. Carpenter described this as a "superpreference" curve, but it is nothing more nor less than the ordering of CHO solutions according to relative palatability as revealed by other methods of testing preference. With NaCl solutions the eight-bottle technique shifted relative intake toward the lower concentrations.

Young and Chaplin (1949) used an eight-bottle technique in presenting NaCl solutions within a self-selected diet that contained many other aqueous solutions. They found maximal intake at .7 NaCl (actually .75 NaCl) which is very near to the known optimal NaCl concentration. With an eight-bottle technique the pattern of intake agreed with the well-authenticated acceptance-rejection function for NaCl solutions, and, incidentally, the pattern of relative acceptability was the same for normal and adrenalectomized rats. The total diet was a factor in their result.

Hence it is clear that a result depends upon whether or not the possibility of choice exists and upon the number of bottles of fluids presented for choice.

Main Characteristics of Our Research Methods

Our recent methods of research have the following characteristics: (*a*) The stimuli have been simple and compound taste solutions. Solutions have a big advantage over natural foods in that precise quantitative control of concentration (hedonic intensity) is possible, and, further, the interactions of solutes with positive and negative valences can be studied. (*b*) The work has been done mainly with need-free (nondeprived) rats. It has been found that need-free subjects are more ready to discriminate among test foods than partly starved, undernourished animals. Further, need-free subjects yield quantitative norms of ingestion and preference that can be used in various studies of dietary depletions, surgical operations, special organic states such as pregnancy, species comparisons, etc. (*c*) Our methods have offered rats a free choice between test fluids that are briefly exposed. Brief-exposure techniques give a cross section of the preferential hierarchy at the time of the test and make it possible to distinguish between such phenomena as acquisition of appetitive and aversive behavior, on the one hand, and approach to organic satiation, on the other. (*d*) The up-and-down psychophysical method yields an objective and quantitative measure of palatability if one uses the sucrose scale as the scale of measure. This method is described below.

PSYCHOPHYSICAL STUDIES OF HEDONIC PROCESSES

The Up-And-Down Psychophysical Method

Prior to introduction of the up-and-down psychophysical method we were limited to preferential hierarchies or transitive series, such as the following (Young, 1949b, p. 40): 50 CHO $>$ 4 CHO $>$ .5 NaCl $>$ H_2O $>$ 3 NaCl. Such series do not reveal the difference in palatability between any two members of the series. The up-and-down method, however, yields an equivalent CHO concentration for each member in the series and thus provides information about the relative spacing of the members along the hedonic continuum.

The up-and-down method was introduced to our laboratory by W. E. Kappauf, beginning with the doctoral research of Christensen (1962) and extending to current studies. The method has been employed in experiments by Kappauf, Burright, and DeMarco (1963), Young and Schulte (1963), and Young and Madsen (1963). These papers should be consulted for technical details. In general, we have followed the computational procedures outlined by Dixon and Massey (1957).

The up-and-down method rests upon the principle that when rats are offered a choice between two sucrose solutions they prefer the higher (sweeter) concentration. The principle applies not only to simple sugar solutions but to compounds that contain sucrose plus one other component such as salt or quinine. The palatability of a simple or compound solution can be raised by increasing, and lowered by decreasing, the sucrose concentration. It should be pointed out, however, that palatability as defined by preference tests is not the same as intake or rate of licking. Maximal intake and maximal activation of the tongue occur at approximately 8.5 CHO.

The up-and-down method was first applied to testing groups of rats. Ordinarily a group of 30 subjects can be tested during a single session. From the data one can compute *for the group as a group* the mean isohedonic concentration, the SD of the sample, and the SE of the mean, as described by Dixon and Massey (1957). The group method yields a measure of hedonic equivalence in a single test. Since the choice of one subject determines the preference test to be given to the next subject, the method reduces to a minimum the effects of experience.

An obvious limitation of the group up-and-down method is that it does not give measures of hedonic equivalence for individual subjects and hence yields no information about individual differences. Young and Madsen (1963), however, demonstrated that the up-and-down method can be applied to testing individuals. Their technique, I believe, is the best available method for testing taste preferences of individual subjects, and I would strongly recommend it to anyone contemplating research in this area. Young and Madsen had no difficulty in obtaining 25 choices per rat in a single session. With individual tests the choice during one trial determines the pair of test fluids to be presented in the next. Their criterion of preference was continuous drinking of one fluid for 5–10 seconds immediately after forced sampling of first one and then the other fluid, followed by a period of free choice.

Palatability Ratings with the Sucrose Scale

Since need-free rats, when offered a choice, prefer the higher (sweeter) concentration of two sucrose solutions, it is possible to use sucrose concentration as a basis for scaling levels of acceptability of fluids and solid foods.

Hedonic scaling is illustrated in Figure 1 which shows the hedonic equivalence of solutions of sodium saccharin and sucrose as determined by Young and Madsen (1963). In collecting the data presented in this curve we selected the seven standard concentrations of sodium saccharin shown along the base line of the figure. For each standard eight rats were tested individually with the up-and-down method to determine the mean sucrose equivalent concentration. For each test the formula was K Sac versus X CHO. The results of a test with X varied up and down established:

$$\mathrm{K\ Sac} \overset{\mathrm{h}}{=} \bar{\mathrm{X}}\ \mathrm{CHO}$$

in which K is constant throughout a test (e.g., .40 Sac), and $\bar{\mathrm{X}}$ is the mean CHO concentration at hedonic equality. The symbol $\overset{\mathrm{h}}{=}$ is here introduced to represent hedonic equality.

Vertical lines in Figure 1 indicate the range of means for the eight subjects. The solid curve shows the mean hedonic-equivalence function for the group as a whole. On the basis of this curve we estimated the optimal concentration at .5 Sac and that this optimum is hedonically equivalent to a

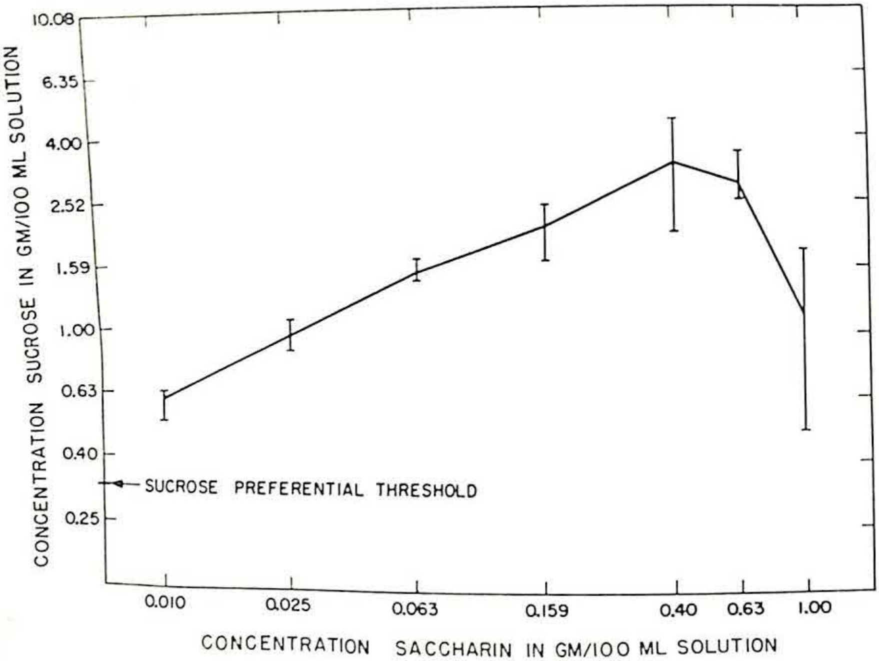

FIG. 1. The hedonic-equivalence function for simple solutions of sodium saccharin and sucrose. (From Young & Madsen, 1963.)

sucrose concentration of 3.5 CHO. Incidentally, human observers report that saccharin solutions are sweet and pleasant up to about .5 Sac and that higher concentrations become sharp (bitter?) and unpleasant.

With suitable modifications of the apparatus it would be possible to obtain objective palatability ratings of solid foods. Young and Greene (1953) ran some tests between solid crystalline sucrose and a solution of 36% sucrose. They found that the solution was preferred by rats to the solid sucrose. We did not use the up-and-down method at the time, but there can be no doubt that preference tests between solids and liquids can easily be made.

Because rats prefer the sweeter solution to the less sweet, we must not jump to the conclusion that a sucrose scale is suitable for hedonic ratings with human subjects. Jerry W. Koppman, in an unpublished study under the direction of W. E. Kappauf, found that some human subjects (like the rats) consistently preferred the sweeter solution, except that 64 CHO (a highly viscous, sticky syrup) was the least acceptable fluid for all subjects. Other subjects tended to prefer the less-sweet solution especially when concentrations were above about 10 CHO. For the group as a whole 16 CHO was the most acceptable concentration, but there were individual differences in the optimum. Koppman concluded that the up-and-down method with a CHO scale is not suitable for palatability ratings with human subjects.

Isohedons

Guilford (1939) proposed that the term *isohedon* be employed to designate a contour line of equal affective value on a stimulus surface such as a triangular surface with white, black, and yellow at the corners. The term was coined by analogy to contour lines of equal barometric pressure (isobars) on weather maps and the like. Subsequently Guilford (1954) published a family of isohedons plotted on a surface defined by the frequency and intensity of auditory stimuli.

With gustatory stimuli a *stimulus area* can be defined by specifying the range of concentrations for two solutes, both of which are present within a compound solution. A stimulus area represents an infinite number of compound solutions containing two solutes. If three or more solutes were present in a compound solution, a tridimensional or multidimensional space would be necessary to represent all possible compounds. To date, however, we have worked only with mixtures containing two solutes.

The following general equation is used in plotting an isohedon, for example, in the CHO–NaCl stimulus area:

$$\text{K CHO} \overset{\text{h}}{=} \bar{\text{X}}\ \text{CHO} + \text{C NaCl}$$

in which K and C are constants throughout a particular test, and $\bar{X}$ is the mean concentration that makes the compound solution and the sucrose standard isohedonic. For example, in determining a point on the 4 CHO isohedon at the level of 1 NaCl we would hold to these values in a test: 4 CHO versus X CHO + 1 NaCl. From 4 to 6 values of X, varied up and down from trial to trial, are ordinarily used in a test. These values are separated by equal log distances and are determined in advance with a group of pilot animals.

Figure 2 is an isohedonic contour map in the CHO–Q stimulus area. It

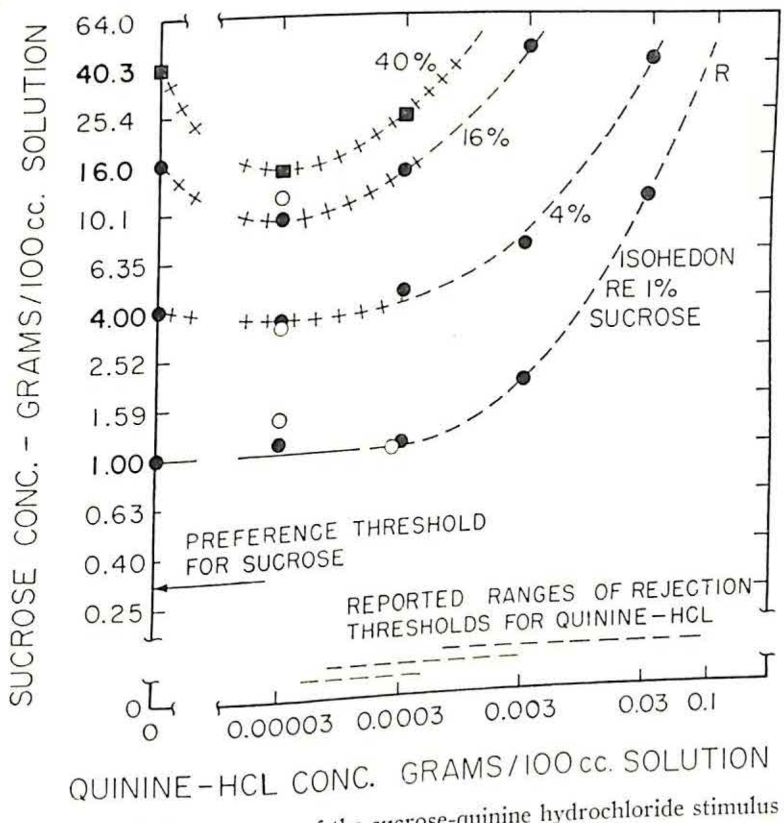

FIG. 2. Isohedonic contour map of the sucrose-quinine hydrochloride stimulus area. (From Kappauf et al., 1963.)

illustrates features common to isohedonic maps in other stimulus areas: (*a*) Most but not all isohedons dip below the level of the standard CHO concentration (shown in Figure 2 by heavy numbers at the left) and also rise above the standard level. In this particular study the dips are indicated by + signs to show that low concentrations of Q enhance the palatability of the mixture, facilitating ingestion, and the tails of the isohedons are indicated by − signs to show that higher concentrations of Q impair palatability, inhibiting ingestion. The + and − signs imply that there is a bipolar continuum extending from positive to negative hedonic values. And, further, the signs show that it is possible to study the combination of solutes with positive and negative valences. (*b*) In the four stimulus areas that have been mapped to date (CHO–NaCl, CHO–Q, CHO–H, CHO–Sac), low concentrations of the nonsucrose variable have a positive or indifferent hedonic influence, facilitating ingestion, and high concentrations have a negative hedonic influence, inhibiting ingestion. (*c*) The isohedons, in general, tend to run similar and roughly parallel courses. They do not intersect each other. This is exactly what we might expect if the curves in a family represent genuine isohedons, since the higher CHO concentrations are known to be preferred by rats to the lower. (*d*) Isohedons leave the stimulus area at relatively high concentrations of the nonsucrose variable. This fact implies that there

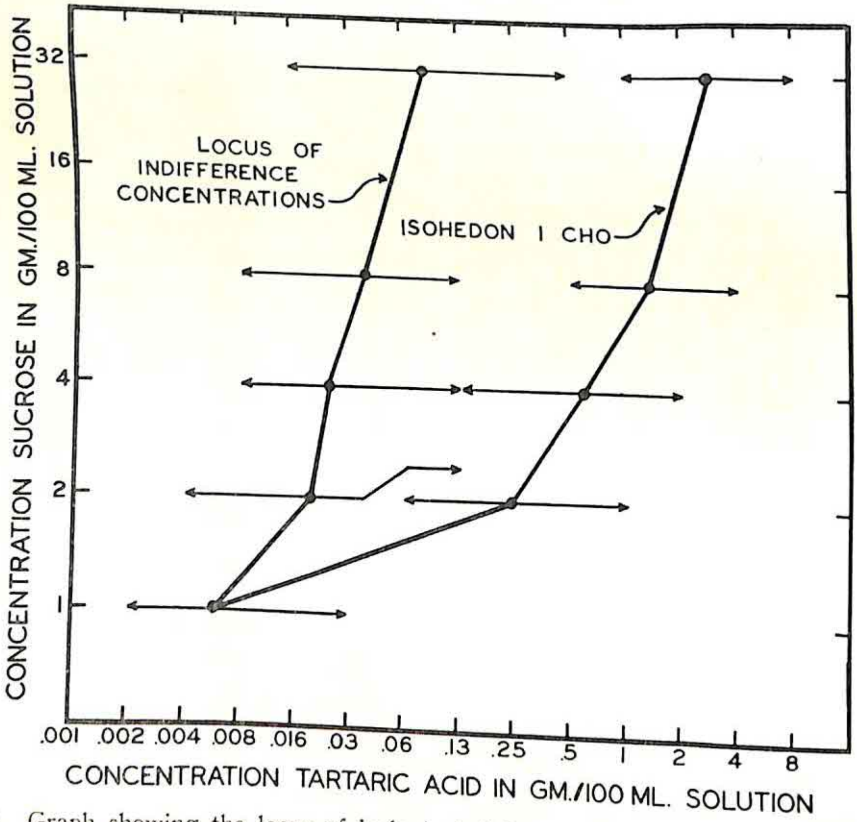

FIG. 3. Graph showing the locus of hedonic indifference (left) and part of Isohedon 1 CHO (right) in the sucrose-tartaric acid stimulus area. (Redrawn from Young & Schulte, 1963.)

is a limiting concentration of the nonsucrose variable at which no amount of sugar in the compound solution can make it acceptable. Rough and preliminary approximations of the upper rejection limits for need-free rats are 6–8 NaCl, .3–.4 Q, 3–4 H, 3 Sac.

The Locus of Hedonic Indifference

If an isohedon dips below the standard level at low concentrations and rises above the standard level at higher concentrations, there is implied an indifferent concentration which neither enhances nor impairs palatability but which has a neutral influence. At the point of indifference the standard and comparison solutions are isohedonic but of different sensory quality.

In a preliminary survey of the sucrose-tartaric acid stimulus area Young and Schulte (1963) determined concentrations of hedonic indifference at five CHO levels. The result is presented in Figure 3. Each indifferent concentration was determined by the test: K CHO versus K CHO + X H. Hedonic equality is represented by the general equation:

$$\text{K CHO} \overset{\mathbf{h}}{=} \text{K CHO} + \bar{\text{X}}\ \text{H}$$

in which K is a sucrose concentration held constant throughout a test and equal in the standard and comparison solutions, and $\bar{\text{X}}$ is the mean acid concentration that makes the comparison solution hedonically equal to the standard.

The horizontal arrows indicate the range of acid concentrations used in each test. For each range all subjects preferred the compound (tart) solution to the standard with concentrations shown at the extreme left; all subjects preferred standard to compound (sour) solution with concentrations at the extreme right. The line connecting the mean indifferent concentrations divides the area into two regions. To the left of the locus of indifference the acid concentrations enhance palatability; to the right of the locus of indifference the higher acid concentrations impair palatability. The line itself represents the locus of compound solutions in which tartaric acid has a neutral hedonic influence.

Every point upon the locus of indifference is also a point upon an isohedon with a standard at the same CHO level. This is illustrated by the curve at the right, Figure 3, which is part of Isohedon 1 CHO. The points on this isohedon were determined by solving for X in this test: 1 CHO versus K CHO + X H. The general equation is:

$$1\ \text{CHO} \overset{\mathbf{h}}{=} \text{K CHO} + \bar{\text{X}}\ \text{H}$$

in which K is the concentration used to determine a point on the locus of indifference, and $\bar{\text{X}}$ is the mean concentration of H at hedonic equality. This is *part* of Isohedon 1 CHO because it does not extend into the region where low concentrations presumably would enhance palatability.

Hedonic Summation and Interaction

If hedonic processes summate algebraically, the addition of NaCl to a CHO solution should enhance palatability when the NaCl concentration is in the range of acceptance and impair palatability when the NaCl concentration is in the range of rejection. Such algebraic summation does, in fact, occur. Young and Christensen (1962) found that the addition of .75 NaCl (near optimal) to a sucrose solution of .5, 1, or 2 CHO made the compound preferred to a simple sucrose standard of the same concentration. And the compounds were also preferred to a simple standard of .75 NaCl. The impairment of palatability by addition of

4 NaCl (hedonically negative) was easily demonstrated.

The summation of positive hedonic intensities did not appear with concentrations of 4 CHO and 8 CHO, but subsequently with another method Christensen (1962) demonstrated hedonic summation at the higher CHO levels. He attributed the success of the second experiment to an improvement of method. Specifically, the improvement provided: (*a*) forced preliminary sampling of both fluids, (*b*) utilization of spatial cues as aids to discrimination, (*c*) the up-and-down psychophysical method.

Algebraic summation of positive hedonic intensities has also been demonstrated in the CHO–Sac stimulus area. Young and Madsen (1963) found that CHO–Sac isohedons made a much bigger dip when .5 Sac was present than isohedons in the CHO–NaCl stimulus area. This phenomenon is understandable as simple algebraic summation since the optimal Sac concentration is equivalent to 3.5 CHO (see Figure 1) while the optimal NaCl concentration is equal to approximately 1 CHO.

The concept of algebraic summation breaks down, however, in the CHO–Q̰ stimulus area. Kappauf et al. (1963) pointed out that neither their own work nor the experiments of Patton and Ruch (1944), Benjamin (1955), and Young, Burright, and Tromater (1963) found any evidençe that simple quinine solutions at any concentration have positive hedonic value. All quinine solutions are rejected at threshold and higher concentrations. Nevertheless, low concentrations of Q combined with high concentrations of CHO produce mixtures that are preferred to simple CHO solutions of equal concentration (see Figure 2). The hypothesis of algebraic summation would require that the addition of a solute with negative valence would impair rather than enhance palatability. Consequently we must postulate that the addition of Q̰ to CHO produces an interaction that differs from simple algebraic summation.

The nature and bodily locus of this interaction are unknown. Interaction could occur on the tongue in the presence of saliva: The ions of a compound solution could have interrelated stimulating effects upon the gustatory receptors and thus influence the total nerve discharge. Or interaction could occur in the nervous system at the midbrain level where primary affective processes appear to originate. The *where* and *how* of hedonic interactions and their relation to simple algebraic summation are matters to be investigated in further experiments.

Diagnostic Tests of Individual Differences in Affective Responses

Figure 4 shows four isohedons in the CHO–NaCl stimulus area with a standard of 1 CHO. The lowest isohedon, plotted with broken lines, is based upon the data of Young and Schulte (1963) for a group of 30 rats tested with the *group* up-and-down method. The other isohedons, plotted with solid lines, show curves based on tests of Young and Madsen (1963) with eight rats tested with the *individual* up-and-down method. The heavy line is the mean isohedon for the eight rats, the light lines show isohedons for two individuals: Rat 3 and Rat 8.

It will be observed that the mean isohedon for the rats tested individually is higher in the field than the isohedon based on data obtained with the group method. Just why this difference exists is not clear, but each point on an individual isohedon is based upon 75 choices per rat (25 per day for 3 days), while each point on the group isohedon

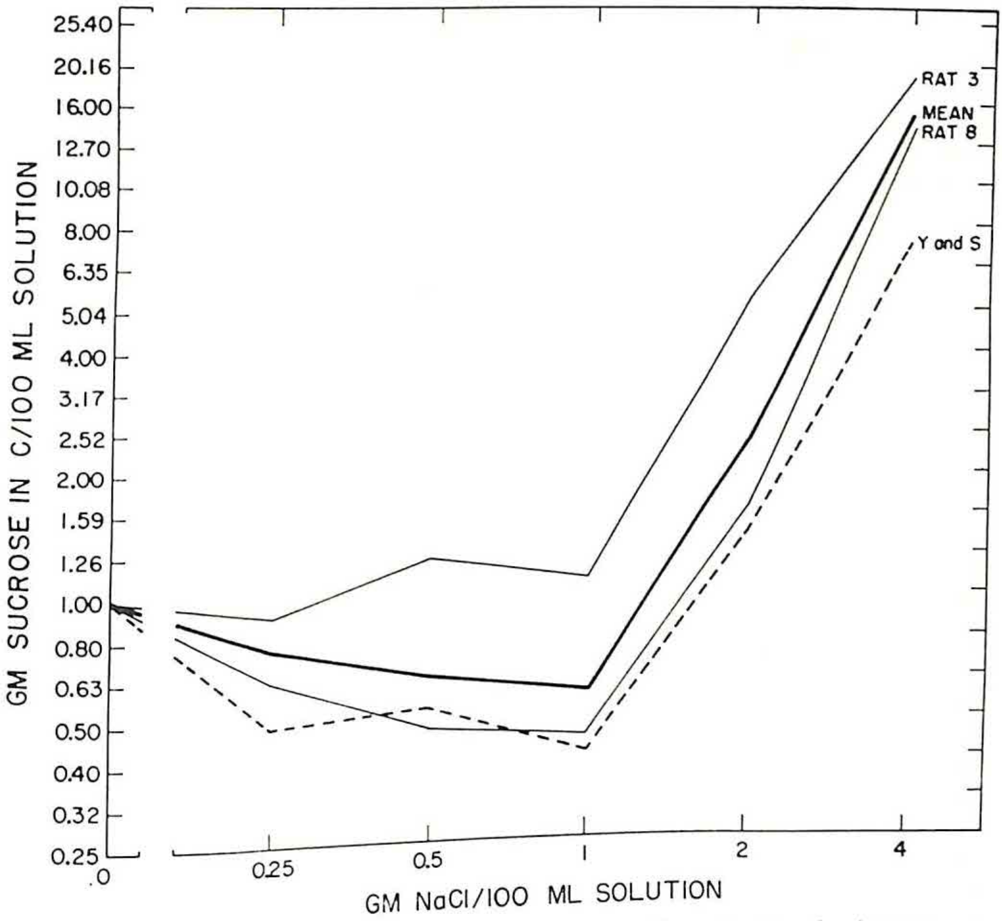

FIG. 4. Group and individual isohedons in the sucrose-sodium chloride stimulus area. (From Young & Madsen, 1963.)

is based upon one choice per rat on a single day. I believe that the greater amount of experience in the individual tests accounts for the difference between the two curves, but further study of this matter is needed.

More interesting is the displacement of curves for individual subjects. The isohedon for Rat 3 is at all points higher in the field than that for Rat 8. Also the curve for Rat 3 does not dip significantly below the 1 CHO standard level while the curve for Rat 8 shows the typical dip. Moreover, we found that in isohedons with standards of 4 CHO and 16 CHO, Rat 3 was consistently the "high" animal and Rat 8 consistently the "low" animal. There is an obvious parallel displacement of isohedons for individual rats. There is, of course, overlap, but curves for some subjects are consistently above the mean, for others, consistently below the mean, for others, consistently average.

This finding of parallel displacement of individual isohedons is theoretically and practically important. It implies that isohedons reveal consistent individual differences in affective responses when tested at any or all points within the stimulus area. Practically this means that in developing diagnostic tests of individual differences in affectivity we do not need to plot total isohedons. Individual differences can be tested at a single point within the stimulus area. Or (even better) diagnostic tests can be made by getting hedonic ratings of simple or compound solutions with a CHO scale. For example, in studying conditions that change the level of acceptability of NaCl one might obtain ratings of 1 NaCl on a sucrose scale, or perhaps ratings of a compound such as 1 NaCl + 4 CHO, and observe how these ratings changed with adrenalectomy, sodium depletion, and other conditions.

The interpretation of individual differences in tests of affective responses is another matter. The difficulty can be illustrated by reference to Figure 4. One can argue that Rat 3 is *less sensitive* than Rat 8 to sucrose stimulations since he requires more CHO in the compound to produce hedonic equality with the standard; Rat 3 has less of a "sweet tooth" than Rat 8. But one can equally well argue that the difference lies in the affective response to NaCl. Rat 3 may react against NaCl more strongly than Rat 8 and hence require more CHO in the compound to counteract a built-in aversiveness. I am sure that this difficulty of interpretation can be overcome by further work and that diagnostic tests of individual differences in affective responses are entirely feasible.

Preference and Activation of the Tongue

Young and Schulte (1963) demonstrated that a distinction must be drawn between preference and activation of the tongue. Their data revealed two interesting facts: First, the amount of licking a physically constant standard solution varied with the comparison solution that happened to be paired with it during a preference test. If the comparison solution was of high palatability, the *standard* solution was licked at a higher rate (licks/minute) than if the comparison solution was of lower palatability. There was thus a gestalt effect that influenced intake. Second, the total amount of licking *both* fluids (standard plus comparison) varied with the location of the comparison solution in the stimulus area. A striking illustration of this was found in the CHO–Q area. With a standard of 1 CHO the mean numbers of licks of *both* fluids (standard plus comparison) are shown in the following tabulation:

Concentration of Q in the comparison compound	.00001	.0001	.001	.01	.1
Mean number of licks with *both* fluids	207	162	141	113	47

It is obvious at a glance that the compound solutions represented by points along an isohedon are *not* equal in their potential for activating the licking response.

Preference is a phenomenon which can occur at any level of tongue activation. Viewed from this angle, an isohedon appears to be a series of points. Each point represents hedonic equality between a *physically* constant standard and a comparison solution tested at a fixed location within a stimulus area.

Activation with Single Gustatory Stimuli

To study tongue activation further Young and Trafton (1964a) recorded rates of licking single solutions (simple and compound) exposed for two successive 60-second periods. They explored the same four stimulus areas that had previously been mapped with isohedons so that activity data with single stimuli could be compared directly with preferences elicited by paired stimuli. Samples of their findings are presented in Figures 5 and 6.

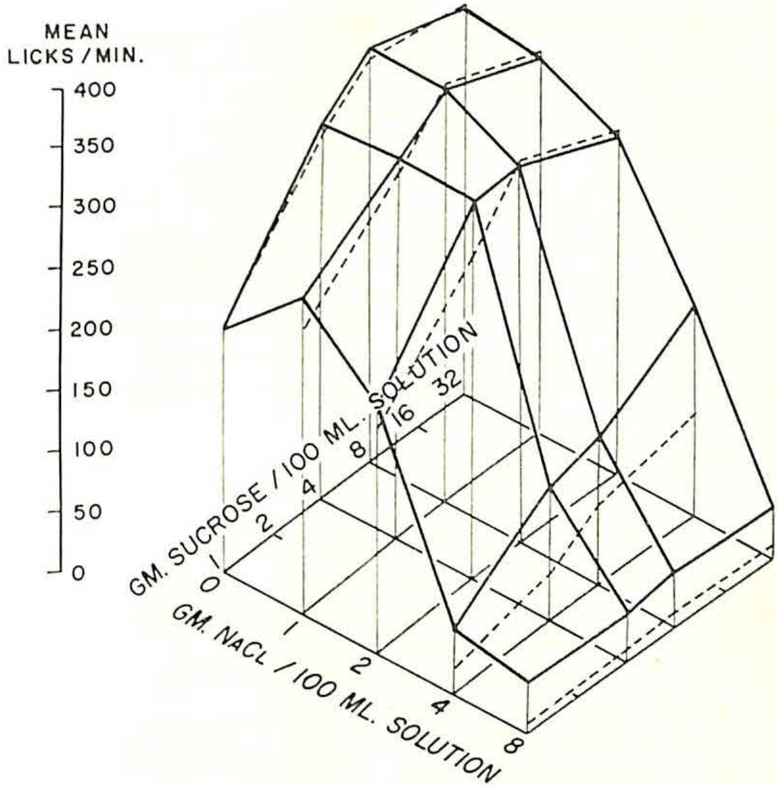

FIG. 5. Tridimensional activity surface for the sucrose-sodium chloride stimulus area. (The means are for a group of eight need-free rats. From Young & Trafton, 1964a.)

Figure 5 shows a typical tridimensional activity surface based upon the mean rates of licking different solutions in the CHO–NaCl stimulus area. Fine vertical lines give rates of licking at the 20 points where tests were made. The solid lines describe a tridimensional surface based upon mean rates

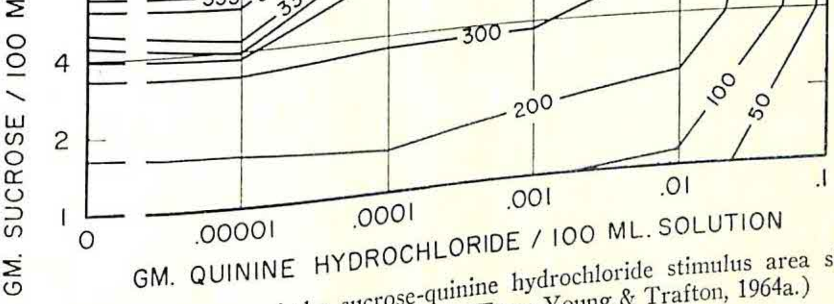

FIG. 6. Activity contour map of the sucrose-quinine hydrochloride stimulus area showing primary and secondary peaks of activity. (From Young & Trafton, 1964a.)

of licking during the first 60-second exposure. Broken lines show rates of licking during a second 60-second exposure that followed after a pause of 30 seconds or more.

Figure 6 is an activity contour map of the CHO–Q stimulus area. Numbers on the contour lines indicate mean numbers of licks per 60 seconds during the first exposure. Each line on the surface is an *isoacton*—a curve representing simple and compound solutions that elicit the same amount of licking.[4]

The following generalizations are based upon a study of tridimensional activity surfaces and activity contour maps in four stimulus areas: (*a*) The highest rates of licking occur where a sucrose concentration of 8 CHO is present alone or in combination with a relatively low concentration of the nonsucrose component. The lowest rates of licking are found where the lowest sucrose concentration (1 CHO) is combined with the highest nonsucrose concentration. (*b*) In all four stimulus areas where the concentration of the nonsucrose component is in the region of negative affectivity (as defined by isohedons and by the loci of indifference) it is possible to find certain concentrations of the nonsucrose component which yield rates of licking having a direct monotonic relation to CHO concentrations in the compound solutions. For example, observe in Figure 5 the relative rates of licking compounds that contain 4 NaCl. (*c*) In the CHO–NaCl and CHO–Sac stimulus areas (but not in the CHO–Q and CHO–H areas) the highest concentrations of the nonsucrose component have a fairly uniform inhibitory influence upon the licking response at all levels of CHO concentration. In Figure 5, for example, compare the activity gradients at 8 NaCl and 4 NaCl. (*d*) In the CHO–NaCl and CHO–Sac stimulus areas there is an indication of algebraic summation of positive hedonic intensities at the level of 1 CHO but not at higher CHO concentrations. Figure 5, for example, shows a higher rate of licking at 1 CHO + 1 NaCl than at 1 CHO + O NaCl. (*e*) In most tests the mean rate of licking is lower during the second 60-second exposure than during the first, but in the CHO–NaCl area there are a few instances of a "warming-up" effect, that is, slightly higher activity during the second exposure than during the first. This effect appears where relatively low NaCl concentrations combine with relatively high CHO concentrations.

Isohedons and Isoactons

Isohedonic contour maps and activity contour maps are totally different kinds of graphic representations. It is impossible to predict the course of isohedons from isoactons and vice versa. Some points of divergence are noted below.

From the activity charts one cannot predict the uniform preference of 32 CHO and 8 CHO. On the basis of amount of licking one would predict little difference in acceptability between these two fluids or that 8 CHO would be preferred to 32 CHO. Young and Trafton (1964a) found that the rate of licking 8 CHO was slightly and consistently higher than the rate of licking 32 CHO, thus confirming Young and Greene (1953). Moreover, the difference in rate of licking cannot be attributed to the approach of satiation since the difference in rate was observed with initial exposure times as brief as 10 seconds.

Again, all activity contour charts (except that for CHO–NaCl) reveal secondary peaks of high activity in places that have no relation to the isohedons. This point is illustrated in Figure 6 which should be compared with the isohedonic contour map of the same area shown in Figure 2. The isohedons indicate a maximal enhancement of palatability at .00003 Q + 40.3 CHO, but the maximal rate of licking is located at about .00001 Q + 8 CHO. Further, the isohedons give no indication of the secondary peak of activity shown in Figure 6 at .001 Q + 8 CHO. Therefore, as noted above, preferential discrimination is one thing, activation of the tongue (intake) is something else.

---

[4] The term *isoacton* was coined by Young and Trafton (1964b) to match Guilford's (1939) term *isohedon*.

It is scarcely necessary to add that the activity charts give no objective basis for distinguishing between positive and negative valences, but this distinction is clearly indicated by isohedons and the loci of hedonic indifference where there is a standard solution. Activity charts do show that increasingly high concentrations of the nonsucrose variable inhibit or retard licking (ingestion) to the point of complete rejection of the compound, but there is no clear line of division between facilitating and inhibiting influences.

Instrumental Activities and Preference

It is easy to assume (as I and others have erroneously done) that the most palatable food is also the best incentive for eliciting instrumental activities. This assumption is true only within limits, and there is a good deal of contradictory evidence: In a bar-pressing experiment Guttman (1953) found that with continuous reinforcement the performance level was higher at 16 CHO than 32 CHO. Young and Shuford (1955) found that the rats' speed of locomotion down a 6-foot runway was faster when the subjects approached an incentive of 18 CHO than one of 54 CHO. Yet there is plenty of evidence that rats prefer the higher CHO concentration to the lower when offered a choice.

In an early experiment (Young, 1944) rats shuttled back and forth on the preference tester with an incentive of crystalline sucrose for one group and whole wheat powder for the other. There were large individual differences in running time, and performance of the groups overlapped so greatly that:

> The safest conclusion to draw is that present results do not afford a sound basis for any prediction of preference between sugar and wheat. The activity of running is one thing and food preference, as tested by our apparatus, is something else [p. 339].

In another experiment, Young, Falk, and Kappauf (1958) measured the time required by rats to approach a cup, 38 centimeters away from the starting box, which contained a solution of .1, .3, .9, 2.7, or 8.1 NaCl. They found that the maximal rate of approach was to the 2.7 NaCl solution. Preference tests had previously shown that optimal palatability was close to .9 NaCl. Hence rate of approach does not correspond exactly with palatability.

In view of these and similar findings one should not expect to predict preferences infallibly from the performance of instrumental activities such as running down a straightaway, bar pressing, T-maze performance, etc., and vice versa.

Studies of Thresholds

Sensory and Preferential Thresholds

Richter (1939), using a continuous-exposure preference technique, reported that the taste threshold of rats for sodium chloride was lowered by adrenalectomy. He believed that the need for NaCl resulting from this operation increased the sensitivity of gustatory receptors.

Richter's (1939) claim instigated a number of experimental studies. One

Young and Shuford suggested that a minimal hedonic intensity is necessary for learning (improvement with repetition) to occur. The matter, obviously, needs further study.

PREFERENCE AND THE REGULATION OF FOOD INTAKE

Regulation of Food Intake by Organic Need States and Drives

Deprivation of a substance necessary to maintain homeostasis frequently (but not invariably) develops a specific hunger, or appetite, for the needed substance. For example, if rats are maintained upon a diet deficient in protein, casein is preferred to sugar; but need-free rats prefer sugar (Young, 1941, 1948). Again, Weiner and Stellar (1951) and Epstein and Stellar (1955) have argued that the *need* for the Na ion in adrenalectomized rats is the major determinant of increased NaCl intake.

Recently Lewis (1960, 1964) showed that appetitive drives produced by adrenalectomy and parathyroidectomy have all the characteristics of general hunger and thirst drives. When adrenalectomized rats were deprived of NaCl, they learned habits of bar pressing and T-maze discrimination for rewards of NaCl solutions. The strength of response varied with the duration of NaCl deprivation. The subjects also discriminated between solid food and NaCl solutions on the basis of their need state.

A specific drive, or appetite, for calcium was produced by parathyroidectomy. This drive had all the characteristics of other appetitive drives. The calcium drive led to formation of instrumental habits that yielded a reward of calcium lactate. The strength of the calcium drive varied with the number of hours of calcium deprivation, and the depleted subjects showed a preference for concentrations that yielded more calcium.

Hence, Lewis concluded, there are specific hunger drives for sodium and calcium, and intake is regulated by metabolic needs.

Optimal Palatability and Excessive Intake

A difficult problem for the homeostatic theory of regulation is to account for excessive intake that occurs when NaCl solutions are presented at optimal concentration. Nelson (1947) reported that when rats had access to a solution of .8 NaCl they drank great quantities of it. Growth was retarded, the kidneys and other organs became enlarged, and there were other symptoms of excessive NaCl intake. The rats took much more NaCl than needed to maintain homeostasis and to remain in sound health when sodium chloride was obtainable at a concentration of .8 NaCl.

This finding has been confirmed by Schmidt (1964) with rats injected with 40 milligrams/kilogram phenobarbital and tested for NaCl acceptance with a one-bottle technique. He found a peak of acceptability with injected animals and also with the placebo controls when the concentration was .9 NaCl. The injected rats:

> drink so much of the solution that they overload the stomach and gut and some of the solution flows out the anus. These rats seem literally unable to withdraw from the drinking tube. Moreover, the anal outflow of solution seems unpleasant to the rat in that drinking is depressed for a couple of days thereafter. Whether it would occur after repeated exposures is not known at this time [p. 204].

Schmidt's (1964) graph clearly shows that the intake of NaCl solutions is a function of concentration; his acceptance-rejection curve is similar to that established by previous workers (see Young & Falk, 1956). Since NaCl solutions at optimal concentration lead to excessive intake of salt and with phenobarbital injections, to excessive drinking, it is clear that internal homeostatic mechanisms do not fully account for the regulation of intake. Intake and taste preference are regulated by different, though interrelated, bodily mechanisms.

Aversive Behavior and Satiation

A distinction must be drawn between (*a*) aversive behavior based upon palatability, and (*b*) the checking of ingestion by internal mechanisms as satiation is approached.

Satiation is a terminal organic state characterized by the fact that the subject, in a constant environment, no longer accepts the kind of food or drink upon which he is satiated. Need-free rats are sometimes incorrectly described as "satiated." If such subjects are offered a solution of sucrose, saccharin, or sodium chloride (which they do not need), they ingest quantities of it approaching the limit of satiation for the particular test food. There are states of partial and total satiation. In partial satiation an animal satiated upon one kind of food will still accept another. In total satiation the subject will not accept any kind of food or liquid.

Distinct from the approach to satiation is the immediate rejection of a food or drink that has a bad taste. For example, if rats are offered a solution of .1 Q, they do not drink it but turn away as soon as contact has been made with the fluid. If they are offered a compound containing Q and CHO, for example, .1 Q + 32 CHO, they may accept it, but with repeated experience they develop an aversion.

Rats *learn* to like and dislike, to accept and reject, to discriminate preferentially, and this *learning* is regulated by the immediate hedonic effect of sensory stimulation.

PHYSIOLOGY OF AFFECTIVE PROCESSES

Receptor stimulation has three main physiological effects. The first is *sensory*. The classical projection systems receiving inputs from various receptors mediate cognitive functions such as localization, discrimination, perception, memory, and thought. The second effect of receptor stimulation is *activating*. The reticular activating system is aroused through collaterals from afferent nerves as well as by inputs from the cerebral cortex. The third effect is *hedonic*. Some arousals by receptor stimulation are positive and facilitating, others are negative and inhibiting, still others are neutral. There are built-in neural mechanisms that mediate primary evaluations in terms of biological "good" or "bad."

Septal Stimulation and Affective Processes

Asdourian (1962) was the first, so far as I know, to combine an approach through experimental hedonism with the self-stimulation technique of James Olds. He attempted to discover whether affective arousals produced by taste stimuli could be combined with self-stimulations of the septal area of the limbic lobe to influence choice. Specifically, he asked: Can self-stimulation through electrodes implanted in the septal area of the limbic system make quinine solutions acceptable? Can normal preferences among sucrose and quinine solutions be modified by brain shock?

He arranged an apparatus in which rats were tested with single-taste fluids or with pairs of fluids presented for choice. The apparatus was so designed that tongue contacts with either or both fluids could produce a brain shock through electrodes implanted within the septal area of the rat's brain.

Prior to tests with taste solutions, Asdourian found that of 26 naïve rats, 12 responded actively by bar pressing that produced septal stimulation, 14 responded so little that they were called "nonresponders" even though they received septal stimulation from bar pressing.

Asdourian found that when tongue contact with a fluid produced brain shock the rate of licking was reduced for all rats from normal (6–8 licks/second) to 2–3 licks/second. Septal stimulation, therefore, retarded licking and ingestion, but it did not inhibit tongue contact with the fluid that produced septal stimulation. The behavior of the subjects gave the impression of "bar pressing with the tongue" to produce septal stimulation! For example, when the rats were offered a choice between 4 CHO without brain shock and 8, 16, or 32 CHO plus brain shock they ingested more of the 4 CHO solution, presumably because the brain shock inhibited ingestion of the higher CHO concentrations rather than because of any change in palatability. The rats made contacts with the fluids that produced septal stimulation but consumed little of them. We might say that the subjects "preferred" septal stimulation to drinking.

Asdourian's data contain evidence that septal stimulation did not disturb the brain mechanisms that normally regulate taste preferences. To be specific: When brain shock was received from contact with *both* fluids presented together for choice, the "responders" showed a normal taste preference of CHO + shock to Q + shock, and the "nonresponders" showed a direct monotonic relation between the numbers of tongue contacts and CHO concentrations of 4, 8, 16, and 32 CHO (see Asdourian's Figure 3). Hence there is evidence that septal stimulations retarded the rate of licking and inhibited ingestion but did not disturb the mechanisms that normally regulate taste preferences.

One subject in the group of "responders" differed from the others in that he was totally indifferent to palatability relations; normal preferences did not influence behavior. When brain shock could be obtained by contact with either fluid, this subject divided contacts evenly between them regardless of palatability thus indicating no preferential discrimination.

Taken as a whole, Asdourian's data indicate that septal stimulation promotes contact with fluids that produce such stimulation but retards normal licking and hence ingestion. One result is certain: We cannot regard septal stimulation as a kind of general pleasantness that can be combined with hedonic effects from tasting quinine, sucrose, or distilled water to change the level of acceptability. Apparently septal stimulation does not affect the neural mechanisms that normally regulate preferences. We conclude that the neural mechanisms that regulate taste preferences are not located within the septal region of the limbic system and that further research will be necessary to determine where these mechanisms are located and how it happens that sweet-tasting fluids normally facilitate, and bitter-tasting fluids normally inhibit, ingestion.

Neural Mechanisms of Appraisal

Arnold (1960) has given a comprehensive definition of emotion which includes complex states such as fear, rage, and lust and also simple feelings of pleasantness and unpleasantness arising from sensory stimulations. Basing her theories on the research of Pribram, MacLean, Papez, and

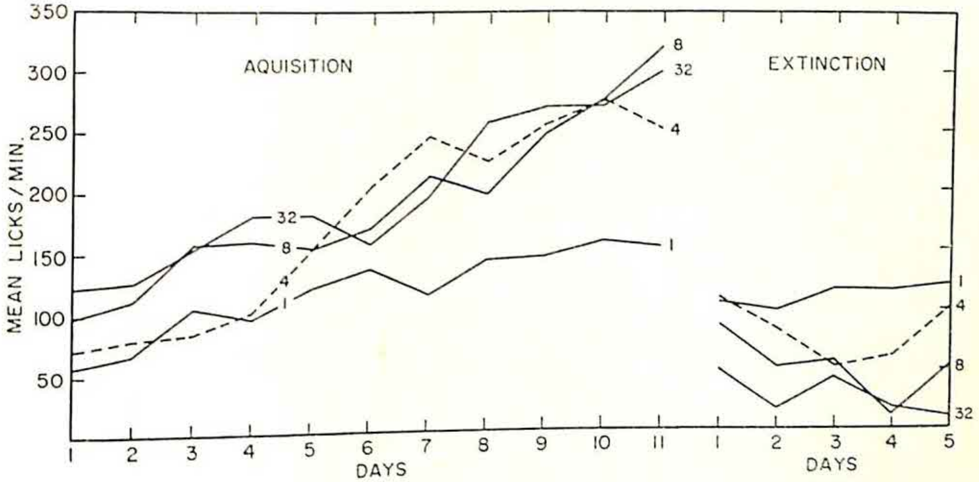

FIG. 7. Acquisition of licking response to four sucrose solutions and extinction with distilled water. (Plotted from unpublished data by Young & Trafton.)

others, she describes an "estimative" neural system which involves the cerebral cortex and subcortical mechanisms. This "estimative" system functions in the appraisal of complex stimulus situations prior to emotional responses. It functions also in the appraisal of simple sensory stimulations.

Arnold regards feelings of pleasantness and unpleasantness as rudimentary appraisals which occur as neural impulses relayed from receptors to the midline and intralaminar thalamic nuclei. It is a matter of theoretical importance that rudimentary appraisals of sensory stimulations are made by built-in neural systems in terms of biological "good" or "bad." Further study of these systems and their analysis at the physiological level is a matter that merits the highest priority.

THE ROLE OF AFFECTIVE PROCESSES IN MOTIVATION AND LEARNING

Acquisition and Extinction of Licking Response to Solutions of Different Hedonic Intensities

The relation between acquisition and extinction of the licking response to sucrose solutions that differ in concentration (hedonic intensity) was explored in an unpublished study by Young and Trafton. The rats, after preliminary exploration of the testing box, were offered a sucrose solution delivered through a nozzle for 60 seconds. The number of licks of the fluid during the exposure was counted electrically. There were four groups of subjects, each composed of eight naïve, need-free rats. The subjects in one group were rewarded with 32 CHO; in another group, 8 CHO; in another, 4 CHO; in another, 1 CHO.

Results are presented in Figure 7. This figure shows that during the first 5 days (apart from two minor inversions on Day 3) the mean numbers of licks per minute were directly proportional to the CHO concentrations. Since the level of performance is proportional to known hedonic intensities, this will be called the *hedonic intensity pattern*. After the first 5 days, however, the curves for Groups 4 CHO, 8 CHO, and 32 CHO cross and recross, and there are no significant differences among them. On Days 10 and 11 the amount of licking was slightly (not significantly) greater with the 8 CHO group than with the 32 CHO group. The rats in the 1 CHO group consistently and significantly licked the fluid at the slowest rate. Since the rates of licking during Days 6–11 agree with what is known about activation of the tongue by CHO solutions (see above discussion), this will be called the *habituated activation pattern*. Learning is marked by a transition from the hedonic intensity pattern to the habituated activation pattern. The influence of hedonic intensity in organizing behavior is thus shown most clearly during the early stages of learning.

After Day 11, all rats were offered an incentive of distilled water for 5 days in the same nozzle. For all groups the substitution of distilled water for a sucrose solution brought an immediate drop in the amount of licking. During the 5 days of extinction, the 1 CHO group showed the *least* change and the 32 CHO group, the *greatest* change in amount of licking. On the fifth day of extinction the amounts of licking were inversely proportional to those of the hedonic-intensity pattern. It is a point of theoretical interest that during extinction the amounts of licking are the inverse of those in the hedonic-intensity pattern rather than the habituated-activation pattern. The explanation, I think, must refer to the acquired disposition which is organized on the basis of hedonic processes.

This finding agrees with results of various other experimenters. For example, Guttman (1953) found a similar inversion of the hedonic-intensity pattern during extinction of bar pressing when the sucrose reward was removed. The biggest drop in rate of bar pressing occurred with rats that had been trained on the highest CHO concentration and the least drop, with rats that had been trained on the lowest concentration. Again, Young and Shuford (1955) observed a similar inversion of the hedonic-intensity pattern during extinction of the habit of running down a 6-foot straightaway for a daily 1-second sip of sucrose solution. Thus the hedonic-intensity pattern is inverted (after learning) when the reward is removed.

Learning Preferential Discriminations with Paired Stimuli

In an early experiment (Young, 1933) abrupt shifts of preference were observed when one test food was suddenly substituted for another. The abrupt shifts were commonly followed by preferential trends which occurred more or less gradually with repeated choices. In subsequent experiments abrupt mutations of preference and more gradual trends have been repeatedly observed and studied. A preferential trend indicates *learning* a preferential discrimination with repeated choices.

To illustrate preferential trends and shifts I have plotted some data (see Figure 8) from a study by Young and Christensen (1962). The curves show percent choices of a compound solution made by a group of eight need-free rats tested on the brief-exposure preference tester. The subjects were offered a series of choices between a simple sucrose standard of 8 CHO or 4 CHO and a compound containing the same concentration of sucrose as the standard plus 2 NaCl or 4 NaCl. In each series of trials the concentration of NaCl in the compound was suddenly increased from 2 NaCl to 4 NaCl.

The curves can be readily interpreted in terms of known hedonic principles. We know that rats prefer 8 CHO to 4 CHO. We also know that a simple 2 NaCl solution is within the range of rejection not far above the indifferent concentration and that 4 NaCl is more

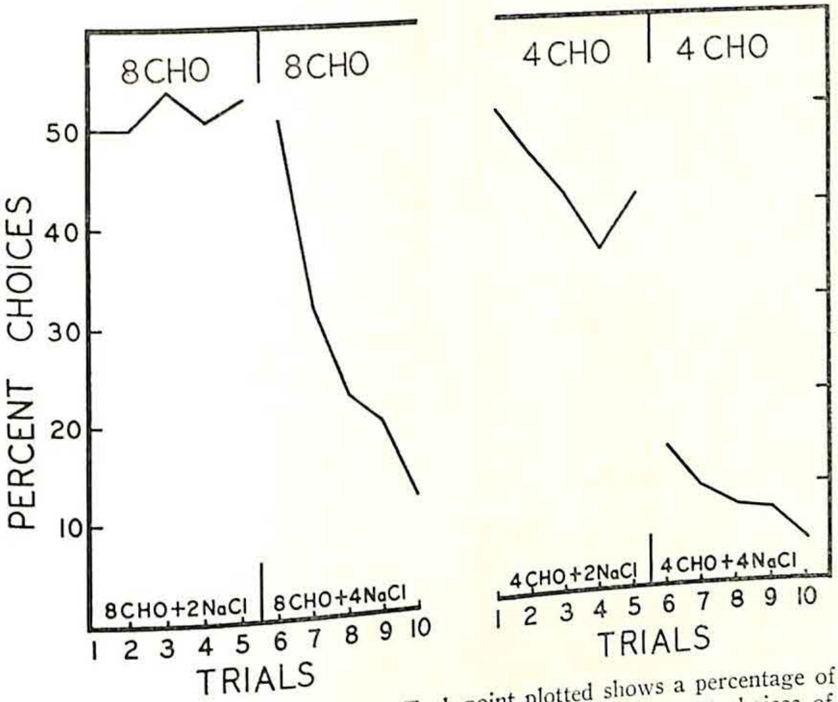

FIG. 8. Preferential trends and shifts. (Each point plotted shows a percentage of choices based on a block of 160 choices—20 choices per rat, 8 subjects. Percent choices of a compound solution containing sucrose and sodium chloride are shown below the curve. Percent choices of the simple sucrose standard, above, equal 100% minus the values plotted. From data by Young & Christensen, 1962.)

intensely negative. Now the first curve at the left is only slightly above the 50% (chance) level. The presence of 2 NaCl in the mixture gave a slight (not significant) rise in the curve, but when the NaCl concentration was suddenly increased from 2 NaCl to 4 NaCl, there was a marked and steady decline in the percent choices of the compound, and the preferential trend moved towards the limit of complete rejection of the compound and complete acceptance of the simple CHO standard. The third curve shows that the addition of 2 NaCl to 4 CHO produced a downward preferential trend rather than an indifferent or weakly postitive effect similar to that which appeared at the 8 CHO level. This is understandable in view of the fact that 4 CHO is less acceptable than 8 CHO and that a weak negative valence may even facilitate acceptance (as with CHO-Q mixture). But when the NaCl concentration was increased from 2 NaCl to 4 NaCl, there was an immediate and abrupt shift in the percent choices of the compound followed by a steady decline—a trend towards complete rejection of the compound and complete acceptance of the simple CHO standard.

These curves indicate immediate preferences. They also show that subjects *learn* to make preferential discriminations. The process of learning does not explain the direction in which a preferential trend moves, the steepness of the slope, nor the abrupt shifts in frequency of choices which appear when incentives are changed, etc. These are *motivational* (nonassociative) phenomena which rest upon known hedonic principles and relationships.

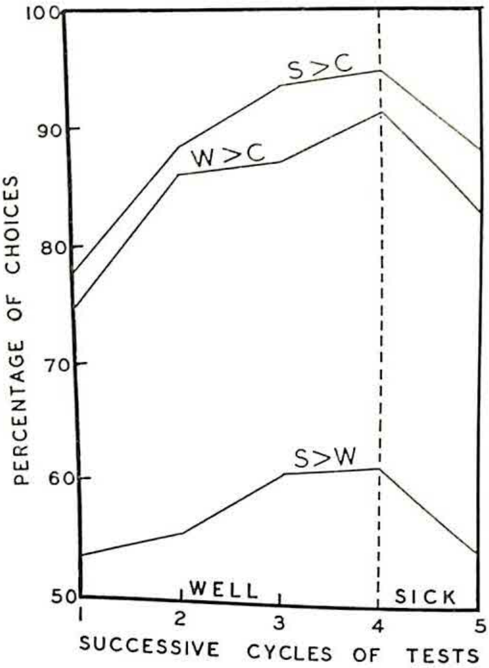

FIG. 9. Development and stability of a preferential hierarchy. (Percentages of choices of the preferred food are shown below a curve, of the nonpreferred food, above. Redrawn from data by Young, 1947.)

Development and Stability of a Preferential Hierarchy

When preference tests are systematically made with three or more test foods, the subjects develop preferential hierarchies which are relatively stable. This came to light in an early experiment (Young, 1947). The study tested three foods of known palatability rating. Crystalline sucrose (S) is preferred by rats to whole wheat powder (W). The difference in palatability between these foods is small, but both are definitely more acceptable than casein (C).

Series of counterbalanced preference tests were run with three pairs of test foods: S-C, W-C, S-W. Cycles of preference tests were planned so that no pair had any advantage of priority, and all single foods had an equal opportunity to influence choice. The results are reproduced in Figure 9.

The curves in Figure 9 illustrate three important points. First, a percentage of choices indicates the *relative* difference in acceptability between a pair of test foods and not absolute hedonic intensity. The lowest curve gives the percentage of choices of the preferred food (S) when paired with W. S and W are high in hedonic intensity (under the diet of the experiment), but the *difference* between them is relatively small as shown by percentages which are near the indiscriminate (50%) level. In the sucrose-casein tests the percentages of choices of sucrose are consistently the highest of all, indicating that the *difference* in palatability between these foods is relatively large. Similarly, in the wheat-casein tests the percentages are almost as high as those in the sucrose-casein tests. This agrees with the fact that W is only slightly less palatable than S.

Second, during the first four cycles all percentages of preference show a simultaneous tendency to rise. The learning of three preferential discriminations went on simultaneously, but the preferential hierarchy did not change. Preferential relations were the same from cycle to cycle. Learning to discriminate did not modify the basic hedonic relations.

Third, during the fifth cycle of tests the rat colony suffered an epizootic of respiratory disease. Several subjects died, others became slow and irregular in running. Figure 9 shows that during this epizootic the subjects were less discriminating than when healthy in the first four cycles. But despite the drops in all percentages and the slower running, the preferential hierarchy remained unchanged.

In general, when rats are systematically tested with three or more foods, they develop a relatively stable preferential hierarchy or value system. This value system is based upon primary hedonic processes which are independent of learning. The acquired value system may persist without change during sickness.

GENERAL CONCLUSIONS

The experiments considered above support the following general conclusions:

1. The postulate that affective processes have an objective existence within the nervous system is confirmed by a group of psychophysical experiments. With the up-and-down psychophysical method affective processes can be studied objectively and quantitatively with remarkable precision. This method has made it possible to (*a*) give objective palatability ratings to simple and compound solutions using rats as subjects and a sucrose scale, (*b*) plot isohedons in four gustatory stimulus areas that are defined by the concentrations of two solutes, (*c*) plot the locus of indifference and thus map regions within a stimulus area where solutes have positive, neutral, and negative hedonic effects, (*d*) provide a method for diagnosing individual differences in affectivity, (*e*) provide norms of acceptability and preference for need-free subjects that can be used in a variety of studies involving deprivation, surgical operation, implanted electrodes, and other conditions.

2. An important distinction has been drawn between activation of the tongue by solutions (intake) and taste preference. The rate of licking (intake) varies systematically with composition of the taste solution when rats are employed as subjects. Tridimensional activity surfaces and bidimensional activity contour maps fail to provide an adequate basis for predicting taste preferences and vice versa. Also, instrumental acts motivated by food reward fail to provide a dependable basis for predicting preferences. Activation is one thing; preference is something else.

3. Some fluids are immediately accepted by the subject, others immediately rejected. If compound fluids that contain solutes of positive and negative valence are accepted, the positive components are prepotent. The use of compound solutions makes it possible to study interaction of positive and negative components within a mixture. Some mixtures show algebraic summation of hedonic intensities, others reveal interaction effects.

4. Measures of preferential taste thresholds vary with the method employed in testing and with motivating conditions. There are different kinds of thresholds: sensory and preferential, physiological and behavioral, thresholds of reinforcement.

5. Current studies confirm earlier findings that the level of acceptability of a food varies with organic need states produced by depletion, surgical operation, and other bodily conditions. Studies of need-free (nondeprived) rats show that the development of aversive behavior through experience (immediate or repeated) must be distinguished from the checking of ingestion by internal organic conditions as satiation is approached. On the positive side there is evidence that acceptance behavior develops on the basis of hedonic effects and that the level of acceptability of a fluid is dependent upon the organic need state.

6. The neurophysiology of pleasantness and unpleasantness is largely unknown and unexplored. The total evidence suggests that sensory stimulations produce primary evaluations, or appraisals, in terms of "good" or "bad" on the basis of innate neural

mechanisms. The genetic basis of primary evaluation is found in untold eons of biological evolution. As a result of evolution there are built-in neural mechanisms that evaluate gustatory stimuli immediately in terms of acceptance-rejection and preference. The nature and functioning of these neural mechanisms can be investigated objectively and quantitatively.

7. A distinction must be drawn between the organizing function of affective processes and the persisting dispositions which are organized. The organizing role of hedonic processes is most clearly apparent during the early stages of development. The acquired dispositions can function as independent organizers and regulators of behavior. In man, the acquired dispositions are called habits, attitudes, expectancies, motives, etc.

8. Some psychologists include the organizing-regulating functions within the concept of motivation, others, within the concept of learning. It is here suggested that hedonic organization and regulation be studied for its own sake and also in relation to motivation and learning. Further, it is suggested that experimental hedonism be extended from rats to other subjects, including the human child and adult, to other kinds of stimulation (especially tactile and thermal), and to other forms of behavior such as sex, play, curiosity, exploration, etc.

REFERENCES

ARNOLD, MAGDA B. *Emotion and personality.* Vol. 2. *Neurological and physiological aspects.* New York: Columbia Univer. Press, 1960.

ASDOURIAN, D. Interaction effects of intracranial stimulation with rewarding and aversive solutions. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 685–690.

BARE, J. K. The specific hunger for sodium chloride in normal and adrenalectomized white rats. *Journal of Comparative and Physiological Psychology,* 1949, **42**, 242–253.

BEEBE-CENTER, J. G., BLACK, P., HOFFMAN, A. C., & WADE, M. Relative per diem consumption as a measure of preference in the rat. *Journal of Comparative and Physiological Psychology,* 1948, **41**, 239–251.

BENJAMIN, R. M. Cortical taste mechanisms studied by two different taste procedures. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 119–122.

BURRIGHT, R. G., & KAPPAUF, W. E. Preference threshold of the white rat for sucrose. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 171–173.

CAMPBELL, B. A. Absolute and relative preference thresholds for hungry and satiated rats. *Journal of Comparative and Physiological Psychology,* 1958, **51**, 795–800.

CARPENTER, J. A. A comparison of stimulus-presentation procedures in taste-preference experiments. *Journal of Comparative and Physiological Psychology,* 1958, **51**, 561–564.

CHRISTENSEN, K. R. Isohedonic contours in the sucrose-sodium chloride area of gustatory stimulation. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 337–341.

DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis.* (2nd ed.) New York: McGraw-Hill, 1957.

EPSTEIN, A. N., & STELLAR, E. The control of salt preference in the adrenalectomized rat. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 167–172.

GUILFORD, J. P. A study of psychodynamics. *Psychometrika,* 1939, **4**, 1–23.

GUILFORD, J. P. System in the relationships of affective value to frequency and intensity of auditory stimuli. *American Journal of Psychology,* 1954, **67**, 691–695.

GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *Journal of Experimental Psychology,* 1953, **46**, 213–224.

IRWIN, F. W. An analysis of the concepts of *discrimination* and *preference.* *American Journal of Psychology,* 1958, **71**, 152–163.

KAPPAUF, W. E., BURRIGHT, R. G., & DEMARCO, W. Sucrose-quinine mixtures which are isohedonic for the rat. *Jour-*

nal of Comparative and Physiological Psychology, 1963, **56**, 138–143.

Koh, S. D., & Teitelbaum, P. Absolute behavioral taste thresholds in the rat. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 223–229.

Lewis, Melissa. Behavior resulting from sodium chloride deprivation in adrenalectomized rats. *Journal of Comparative and Physiological Psychology*, 1960, **53**, 464–467.

Lewis, Melissa. Behavior resulting from calcium deprivation in parathyroidectomized rats. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 348–352.

Nelson, D. Do rats select more sodium chloride than they need? *Federate Proceedings of the American Physiological Society*, 1947, **6**(1, Whole No. 169).

Patton, H. D., & Ruch, T. C. Preference thresholds for quinine hydrochloride in chimpanzee, monkey, and rat. *Journal of Comparative Psychology*, 1944, **37**, 35–49.

Pfaffmann, C. The pleasures of sensation. *Psychological Review*, 1960, **67**, 253–268.

Pfaffmann, C. Taste, its sensory and motivating properties. *American Scientist*, 1964, **52**, 187–206.

Pfaffman, C., & Bare, J. K. Gustatory nerve discharges in normal and adrenalectomized rats. *Journal of Comparative and Physiological Psychology*, 1950, **43**, 320–324.

Pfaffman, C., Young, P. T., Dethier, V. G., Richter, C. P., & Stellar, E. The preparation of solutions for research in chemoreception and food acceptance. *Journal of Comparative and Physiological Psychology*, 1954, **47**, 93–96.

Richter, C. P. Salt taste thresholds of normal and adrenalectomized rats. *Endocrinology*, 1939, **24**, 367–371.

Richter, C. P., & Campbell, K. H. Sucrose taste thresholds of rats and humans. *American Journal of Physiology*, 1940, **128**, 291–297.

Schmidt, H. Water intake as an index of drug action. In M. J. Wayner (Ed.), *Thirst, Proceedings of the First International Symposium on Thirst in the Regulation of Body Water Held at the Florida State University in Tallahasse, May 1963.* Oxford: Pergamon Press, 1964. Pp. 185–209.

Weiner, I. H., & Stellar, E. Salt preference of the rat determined by a single-stimulus method. *Journal of Comparative and Physiological Psychology*, 1951, **44**, 394–401.

Young, P. T. Relative food preferences of the white rat: II. *Journal of Comparative Psychology*, 1933, **15**, 149–165.

Young, P. T. The experimental analysis of appetite. *Psychological Bulletin*, 1941, **38**, 129–164.

Young, P. T. Studies of food preference, appetite and dietary habit: I. Running activity and dietary habit of the rat in relation to food preference. *Journal of Comparative Psychology*, 1944, **37**, 327–370.

Young, P. T. Studies of food preference, appetite and dietary habit: V. Techniques for testing food preference and the significance of results obtained with different methods. *Comparative Psychology Monographs*, 1945, **19**(1, Whole No. 98).

Young, P. T. Studies of food preference, appetite and dietary habit: VII. Palatability in relation to learning and performance. *Journal of Comparative and Physiological Psychology*, 1947, **40**, 37–72.

Young, P. T. Appetite, palatability and feeding habit: A critical review. *Psychological Bulletin*, 1948, **45**, 289–320.

Young, P. T. Food-seeking drive, affective process, and learning. *Psychological Review*, 1949, **56**, 98–121. (a)

Young, P. T. Studies of food preference, appetite and dietary habit: IX. Palatability versus appetite as determinants of the critical concentrations of sucrose and sodium chloride. *Comparative Psychology Monographs*, 1949, **19**(5, Whole No. 102). (b)

Young, P. T. The role of hedonic processes in the organization of behavior. *Psychological Review*, 1952, **59**, 249–262.

Young, P. T. The role of hedonic processes in motivation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation, 1955.* Lincoln: Univer. Nebraska Press, 1955.

Young, P. T. Psychologic factors regulating the feeding process. *American Journal of Clinical Nutrition*, 1957, **5**, 154–161.

Young, P. T. The role of affective processes in learning and motivation. *Psychological Review*, 1959, **66**, 104–125.

Young, P. T., Burright, R. G., & Tromater, L. J. Preferences of the white rat for solutions of sucrose and quinine hydrochloride. *American Journal of Psychology*, 1963, **76**, 205–217.

Young, P. T., & Chaplin, J. P. Studies of food preference, appetite and dietary habit: III. Palatability and appetite in relation to bodily need. *Comparative Psychology Monographs,* 1945, **18** (3, Whole No. 95).

Young, P. T., & Chaplin, J. P. Studies of food preference, appetite and dietary habit: X. Preferences of adrenalectomized rats for salt solutions of different concentrations. *Comparative Psychology Monographs,* 1949, **19** (5, Whole No. 102).

Young, P. T., & Christensen, K. R. Algebraic summation of hedonic processes. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 332–336.

Young, P. T., & Falk, J. L. The relative acceptability of sodium chloride solutions as a function of concentration and water need. *Journal of Comparative and Physiological Psychology,* 1956, **49**, 569–575.

Young, P. T., Falk, J. L., & Kappauf, W. E. Running activity and preference as related to concentration of sodium-chloride solutions. *American Journal of Psychology,* 1958, **71**, 255–262.

Young, P. T., & Greene, J. T. Quantity of food ingested as a measure of relative acceptability. *Journal of Comparative and Physiological Psychology,* 1953, **46**, 288–294.

Young, P. T., & Madsen, C. H., Jr. Individual isohedons in sucrose-sodium chloride and sucrose-saccharin gustatory areas. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 903–909.

Young, P. T., & Schulte, R. H. Isohedonic contours and tongue activity in three gustatory areas of the rat. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 465–475.

Young, P. T., & Shuford, E. H., Jr. Quantitative control of motivation through sucrose solutions of different concentrations. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 114–118.

Young, P. T., & Trafton, C. L. Activity contour maps as related to preference in four gustatory stimulus areas of the rat. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 68–75. (a)

Young, P. T., & Trafton, C. L. Psychophysical studies of taste preference and fluid intake. In M. J. Wayner (Ed.), *Thirst, Proceedings of the First International Symposium on Thrist in the Regulation of Body Water Held at the Florida State University in Tallahassee, May 1963.* Oxford: Pergamon Press, 1964. Pp. 271–285. (b)

(Received September 21, 1964)

Psychological Review
1966, Vol. 73, No. 1, 87-103

ANALYSIS OF THE OVERLEARNING REVERSAL EFFECT[1]

ELIJAH LOVEJOY

University of Pennsylvania

The Overlearning Reversal Effect (ORE) is often found in visual discrimination problems and seldom found in simple position problems. It is argued that the ORE is found only when the animal's initial probability of attending to the relevant stimuli is low. In such a problem, nonovertrained animals may stop attending to the relevant stimuli during reversal and so will learn the reversal slowly. 2 mathematical models support the plausibility of the above notions and confirm the usefulness of "attention" as a variable in discrimination learning theory. Some exceptions to the general rule at the top of this paragraph are reinterpreted.

Before we inquire into possible "explanations" of the Overlearning Reversal Effect (ORE), it will be well to review briefly the relevant experimental results against which any theory must be measured. In 1953, Reid reported that overtrained rats could reverse a black-white discrimination habit more quickly than animals who had been trained to a criterion but not overtrained. This surprising phenomenon has come to be known as the Overlearning Reversal Effect. Reid's result was confirmed by Capaldi and Stevenson (1957), Komaki (1961), Mackintosh (1962, 1963b, 1963c), and Pubols (1956, Experiment 1). All the above experiments involved some variant of a black-white discrimination problem. At the same time, the effect has not always been found in the black-white problem. After finding a marginally significant ORE in 1961 (see D'Amato & Jagoda, 1961), D'Amato and his co-workers have repeatedly been unable to demonstrate the effect when studying brightness discrimination in their automatic Y mazes (D'Amato & Schiff, 1965). Moreover Erlebacher (1963) did a large and careful study using a conventional T maze and found no effect. It seems that in brightness problems the ORE usually, but not always, appears.

A number of studies have investigated the effect of overtraining on the reversal of a simple position habit. In general it is found either that overtraining has no effect on reversal trials to criterion (Clayton, 1963b; D'Amato & Jagoda, 1962; D'Amato & Schiff, 1964, Experiment 2; Hill, Spear, & Clayton, 1962, Experiment II; Komaki, 1962; Mackintosh, 1963a, Experiment V, Group I) or that it produces a significant decrement in speed of reversal learning (Clayton, 1963a; Galanter & Bush, 1959; Hill & Spear, 1963; Hill, Spear, & Clayton, 1962, Experiments I, III, and IV; Kendler & Kimm, 1964; Mackintosh, 1963a, Experiment V, Group II). Again there are exceptions: Pubols in 1956 found the ORE in a position problem, and so did Capaldi in 1963. We will consider these "exceptions" in more detail later. For the moment it suffices to note that the ORE is often found in brightness problems, and rarely found in simple position learning. How are learning theorists to reconcile this dichotomy, and how are they to explain the ORE?

---

[1] This work was supported in part by Grant NSF-GB-1855 from the National Science Foundation to the University of Pennsylvania. The author is grateful to Robert R. Bush for his advice, guidance, and encouragement while this work was being carried out.

First, let us consider why the ORE should be surprising. Reinforcement learning theorists have generally made the following assumptions: (*a*) Habit strength increases monotonically with the number of reinforcements; (*b*) the greater the habit strength, the greater the resistance to extinction; (*c*) reversal learning is simply a matter of extinguishing an old habit and then learning the opposite one. From these three assumptions, theorists have properly deduced that the ORE should not occur. Since it does occur, one or more of the three assumptions must be unjustified. The goal of this paper is to consider which of these three should be changed and to propose a modification.

To the writer's knowledge, the ORE has not prompted anyone to suggest that we abandon the first assumption. A number of proposed explanations would change the second assumption, and even more would modify the third. As for the second, there is some evidence that resistance to extinction is not a monotone function of the number of rewards, even under 100% reinforcement (e.g., North & Stimmel, 1960; Siegel & Wagner, 1963). But it is hard to see how the second assumption can be modified so that it predicts the ORE in brightness problems, but not in position learning. Moreover, there are data within the very experiments in which an ORE is found which suggest that resistance to extinction is greater for the overtrained subjects than for the nonovertrained ones—in accordance with the second assumption. The length of the initial string of errors in reversal seems a reasonable indicator of resistance to extinction. And there is substantial evidence that this string is longer for overtrained than for nonovertrained subjects (Hill et al., 1962; Mackintosh, 1962). Furthermore, in the one test which has been made of trials to extinction with and without overtraining (Mackintosh, 1963c), it was found that the overtrained subjects took longer to extinguish.

Thus it appears that in order to fit the ORE into learning theory we will have to change assumption (*c*). We cannot maintain that reversal learning is simply a matter of extinguishing an old habit and then learning the opposite one.

Explanations for the ORE which would revise the third assumption generally have assumed that for some reason there is positive transfer from original learning to reversal learning; that is, during overtraining the rats learn something which is useful to them in reversal. Such explanations include Reid's "response of discriminating" (1953), Spence's "reduction of position-response tendencies" (as quoted in Reid, 1953), Harlow's "reduction of error factors" (1949), and Sutherland's "switching in of stimulus analysers" (1959). These last explanations vary in the degree to which they are specific to the problem at hand. The most general proposal might be of the "learning set" type, and assume that all learning experience is helpful for all future problem solving. While this could clearly predict an ORE, it could not easily predict that an ORE will be found in a brightness situation and not in position problems.

The theory that will be developed in this paper, although developed independently, is quite close to Sutherland's idea of switching in stimulus analyzers, and to the thinking of Zeaman and House (1963). Those familiar with Wyckoff's (1952, 1954) papers will notice a formal similarity between his ideas and those presented here.

It is argued that in order to solve a discrimination problem, animals must first learn to attend to the relevant dimension (e.g., brightness), and then they must learn to make the appropriate response once they have attended to that dimension. One rationale for such a theory follows: An animal in a two-choice apparatus like a T or Y maze makes a "decision" at the choice point. He then proceeds to the end of the alley and is either rewarded or nonrewarded. At the time of reward or nonreward, the rat must have stored information about which choice he made if he is ever to "learn" from the experience. (This information may be in memory, or it may be available externally in goal-box cues.) For if the rat in the goal box does not "know" whether he chose white or black, then he cannot associate either of these stimuli with reward or nonreward. Now surely a rat has limited information-storage capacity. On a given trial he might remember which side he chose, whether he approached or avoided some special odor, whether he repeated his previous response, whether he went to black or white, whether he went to the side where he was most recently rewarded, etc. But he probably will not remember all these various attributes of the choice he made. The development which follows is based on the simple idea that a rat can attend to only some of the possible dimensions on any given trial, and that in some sense *learning* can occur only on those trials when the subject is attending to the relevant dimension. It has been argued before that such a "selective attention" mechanism could explain the ORE (Mackintosh, 1962; Sutherland, 1959). The argument is that overtrained animals will continue to attend to the relevant dimension throughout reversal and so will learn the reversal quickly. Nonovertrained animals, on the other hand, will abandon the relevant dimension after reversal because it is consistently leading them to the "wrong" choice. Only after a period of responding to other dimensions will the animals rediscover the relevant dimension and learn the reversal. This seems a reasonable sort of explanation, but it is somewhat loose. In the sequel, two different mathematical models are presented which are based on this type of verbal "attention" theory. It will be seen that although they differ slightly in the assumptions made, both predict the ORE in some cases and not in others. For both these models, the correct choice *in initial learning* will be denoted by W and the incorrect choice, by B. This simplifies discussion of reversal learning and eliminates the necessity of phrases like "the new correct choice for reversal, which was previously the incorrect choice . . . ." This notation should not be interpreted to mean that these models are concerned only with "brightness" problems. As will be seen, they are fairly general.

MODEL I—A TWO-STAGE MODEL

Assumptions

The first attention-motivated model which will be presented is a two-stage chained model. It assumes that on each trial an animal may or may not attend to the relevant stimulus dimension, with the probability of attending on Trial n denoted by $P_n(A)$. If he attends, he then has a certain conditional probability of making the correct choice, $P_n(W|A)$. If he fails

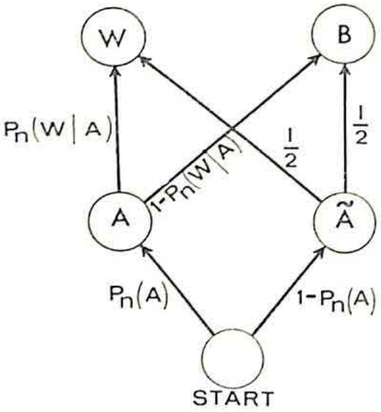

FIG. 1. The state diagram for the first model. (On any trial a subject's performance is represented by one of the four paths: AW, AB, $\tilde{A}W$, and $\tilde{A}B$.)

to attend, we will assume that he has a success probability of $\frac{1}{2}$. Thus four different events are possible:

1. The animal may attend and choose white: Event AW (rewarded).
2. The animal may attend and choose black: Event AB (not rewarded).
3. The animal may fail to attend and go to white: Event $\tilde{A}W$ (rewarded).
4. The animal may fail to attend and go to black: Event $\tilde{A}B$ (not rewarded).

This model is portrayed in the tree in Figure 1.

We must now make some assumptions about how the various events change the two probabilities $P(A)$ and $P(W|A)$. For this model we will assume that reward and nonreward affect the attention process just as they affect overt responses. That is, we will assume that when Event AW occurs (the animal attends, chooses white, and is rewarded), then on the next trial he will be more likely to attend, and having attended he will be more likely to make the correct choice. Both $P(A)$ and $P(W|A)$ will be increased, each approaching a limiting value of unity. When Event AB occurs (the animal attends, chooses black, and is not rewarded), then, since attending was not rewarded, it is assumed that $P(A)$ will decrease; but going to *black* after attending was not rewarded, so it is assumed that $P(W|A)$ will increase. *When the animal does not attend, it is assumed that no changes will be made in* $P(W|A)$. This is an important assumption in the present model and differs from the corresponding one made in the second model. For the present model, when Event $\tilde{A}W$ occurs, not attending is rewarded, and so $P(A)$ will decrease. But $P(W|A)$ will not change. Similarly, when Event $\tilde{A}B$ occurs, not attending is nonrewarded, so $P(A)$ should increase, while $P(W|A)$ is unchanged. These results are summarized in Table 1.

We now have the basic framework for a model. We must still assume the exact amount of the trial-by-trial changes in the two probabilities. In hopes of mathematical simplicity, linear transformations are assumed. It should be mentioned that this is a most arbitrary assumption and is made strictly for mathematical convenience. It is quite possible that some other probability transformation rule, while still maintaining the limit points implied by Table 1, could provide an even better description of the real world. Linear operators

TABLE 1

CHANGES IN $P(A)$ AND $P(W|A)$ WHEN THE FOUR DIFFERENT EVENTS OCCUR

Event	Change in $P(A)$	Change in $P(W\|A)$
AW (rewarded)	+	+
AB (nonrewarded)	−	+
$\tilde{A}W$ (rewarded)	−	0
$\tilde{A}B$ (nonrewarded)	+	0

which will change the two probabilities in the directions indicated in Table 1 are written in Table 2.

At this point Model I has been completely specified. It would be nice to derive some properties of the model, such as the explicit formula for the mean learning curve, the expected total number of errors, etc. This has not been possible, despite considerable effort on the part of the writer and others. However, we can get some idea about the properties of the model by the "stat-rat" method, discussed in detail by Bush and Mosteller (1955). In brief, this method produces sample response protocols of the type that would be found in a population of "statistical rats" which behaved according to the assumptions above. Before we look at some stat-rat computations, however, let us consider at an intuitive level the conditions which might produce an ORE with the present model.

Consider an animal who starts out with an initial probability of attending to the relevant dimension, $P_1(A)=.5$. Suppose he has no response bias, so that $P_1(W|A)=.5$, too. Now in order to solve the problem, the rat must learn to attend on all trials and to always go to white once he has attended. Looking back at Table 1, we can see that $P(W|A)$ cannot decrease. It can only increase or remain unchanged. The probability of attending, on the other hand, may increase or decrease, depending on the event which occurs. Thus if similar rate parameters govern the changes in both probabilities, we would expect $P(W|A)$ to increase faster than $P(A)$. Indeed, when one looks at stat-rat curves, this is what happens. $P(W|A)$ reaches asymptote well before $P(A)$ does. Because of this, animals who have just met a criterion of learning but have not been overtrained have $P(W|A)$ about equal to 1, and $P(A)$ lower, perhaps about .9. Overtrained animals have both probabilities very close to 1. At the start of reversal, then, the nonovertrained subjects have a very low probability of Event AB—since $P(B|A) = 1-P(W|A) \doteq 0$. The probability of Event $\tilde{A}B$, while small, is considerably greater. Thus these subjects are likely to get their first reversal reinforcement by Event $\tilde{A}B$. But this will reduce their probability of attending, since they have been rewarded for not attending, and Event $\tilde{A}B$ will become even more likely. As this sort of process recurs, the nonovertrained subjects will be likely to stop attending to the relevant dimension—and hence to stop learning the reversal. The overtrained subjects, on the other hand, have both $P(A)$ and $P(W|A)$ very close to 1 at reversal. And if the two nonreward rate parameters, α_2 and α_4, are approximately equal, then $P(A)$ and $P(W|A)$ will decrease together during the string of initial reversal errors. For these subjects, $P(AB)$ will be greater than $P(\tilde{A}B)$ early in reversal. Thus they will be more likely to get their first reversal reinforcement while attending. This, in turn, will increase $P(A)$, decrease $P(W|A)$, and increase the likelihood that future re-

TABLE 2

LINEAR TRANSFORMATIONS OF THE PROBABILITIES $P_n(A)$ AND $P_n(W|A)$ WHICH ARE POSTULATED TO OCCUR AFTER THE VARIOUS EVENTS

Event	$P_{n+1}(A)$	$P_{n+1}(W\|A)$
AW	$\alpha_1 P_n(A)+1-\alpha_1$	$\alpha_3 P_n(W\|A)+1-\alpha_3$
AB	$\alpha_2 P_n(A)$	$\alpha_4 P_n(W\|A)+1-\alpha_4$
$\tilde{A}W$	$\alpha_1 P_n(A)$	$P_n(W\|A)$
$\tilde{A}B$	$\alpha_2 P_n(A)+1-\alpha_2$	$P_n(W\|A)$

Note.—α_1, α_2, α_3, and α_4 are rate parameters such that $0 < \alpha_i < 1$.

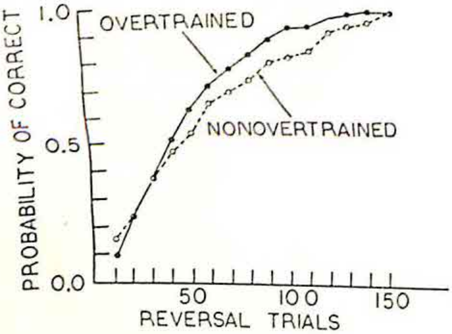

FIG. 2. Group reversal curves for two groups of stat rats, calculated according to Model I. (For original—prereversal—learning it was assumed that $P_1(A) = .5$.)

wards will come by AB rather than $\tilde{A}B$. These subjects will be likely to keep $P(A)$ high during reversal. So they can decrease $P(W|A)$ by means of reversal operators analogous to those in Table 2 and learn the reversal comparatively quickly.

The essence of this argument is that overtrained animals will continue to attend throughout reversal, while nonovertrained ones will be likely to stop attending to the relevant dimension for a considerable length of time during reversal. This will produce the ORE.

Predictions

In fact, if stat rats are run with conditions like those above, an ORE is found. In Figure 2 are plotted the *reversal* learning curves for two groups of stat rats. These statistical animals started original learning with $P_1(A) = .5$ and $P_1(W|A) = .5$. It was assumed that the same pair of rate parameters governed the changes for both the probabilities, so that $\alpha_1 = \alpha_3 = .93$ (the reward parameter) and $\alpha_2 = \alpha_4 = .99$ (the nonreward parameter; see Table 2). A high value of α in the operators shown in Table 2 corresponds to a *slow* change in probability. Thus nonreward was here assumed to be a less powerful event than reward. This asymmetry is desirable because of the data found in such experiments. Just after reversal, overtrained animals may make as many as 40 errors before the first reversal success (the average found by Mackintosh, 1962). This implies a very weak effect of nonreward. But learning can take place fairly quickly in the black-white problem; so the reward operator must be more powerful, as was assumed here.

Animals in the nonovertrained group in Figure 2 were trained in original learning until they had 15 consecutive trials without error. At this point they were reversed, and the group reversal learning curve is plotted. The overtrained animals were trained for 150 trials past this criterion (i.e., they were trained well past asymptote) and were then reversed. Their group learning curve is also shown. Plainly, the overtrained group starts with a lower initial probability of correct, but after a small amount of reversal training the curves cross, and the overtrained group stays superior from that point on. The present model has predicted an ORE.

Experimenters reporting the ORE have not presented learning curves, however. They have simply reported the trials to reach criterion in reversal for overtrained and nonovertrained animals. As D'Amato and Schiff (1965), have pointed out a crossover in learning curves may occur so late in reversal that no significant difference is found in reversal trials to criterion. For the stat rats reported above, however, the crossover occurs early, and there is a significant ORE in terms of trials to criterion. The nonovertrained subjects required 108.3 trials to reverse and the overtrained, 86.1, on the average. These figures may be compared with acquisition where stat rats required 77.2 trials to reach the same criterion. These are the means for groups of 60 animals, and a *t* test applied to the difference indicates significance at the .005 level.

Those familiar with previous experiments on the ORE may wish to know more than the means and the group learning curves reported above. Do overtrained animals uniformly reverse more quickly than nonovertrained subjects, or is the difference merely one of averages? This question is answered in Figure 3. Figure 3 shows the empirical frequency distribution of trials required to complete reversal criterion for the two groups of stat rats. It may be seen that the overtrained subjects *tend* to learn the reversal more quickly, but it is clear that there is substantial overlap between the two distributions. The groups are different, but the reversal score for a sample subject would not suffice for identifying his previous treatment.

Part of our goal has been accomplished. We have shown that a simple set of assumptions, motivated by the idea that attention is important in discrimination learning, have produced a model in which the ORE can occur. But there remains an important step to be taken. Recall that the ORE is generally *not* found in position problems. We must now consider the conditions under which the present model predicts no ORE.

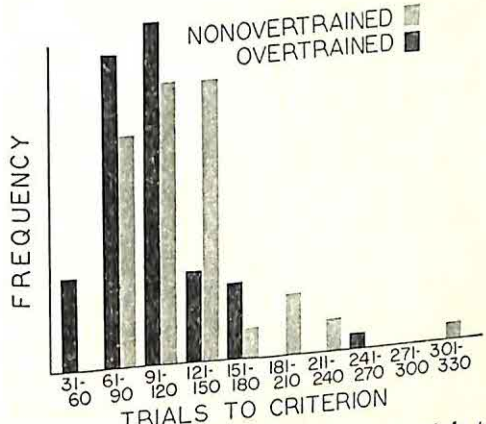

FIG. 3. Empirical distributions of trials to complete reversal criterion for groups of 60 stat rats reversed with and without overtraining. (The overtrained animals tend to learn more quickly.)

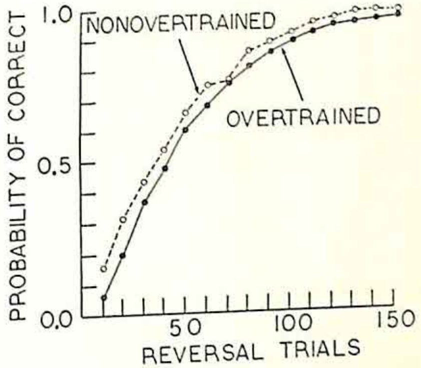

FIG. 4. Group reversal curves for two groups of stat rats calculated according to Model I. (These groups are comparable to those in Figure 2 in every detail, except that for these animals it was assumed that $P_1(A) = 1.0$. That is, the stimuli were assumed very obvious in original learning.)

In discussing the intuitive reasons why the present model should predict an ORE, we pointed out that for the nonovertrained animals $P(W|A)$ reached asymptote before $P(A)$, since both started low and $P(W|A)$ increased faster. But the situation is different when the initial probability of attending, $P_1(A)$, is high. This might well be the case for a position problem. The rat is a nocturnal animal, and it seems reasonable to assume that position is a very obvious cue for such a beast. If we assume that it is completely obvious, that is, that $P_1(A) = 1$ for the rat in a position problem, then $P(A)$ starts out at asymptote. So unless it were to drop sharply—and there is no reason to expect that it would—then no "first reversal reinforcement" argument like the one for brightness problems can be made. We have no reason to expect the ORE.

In Figure 4 are shown reversal learning curves for two groups of stat rats. The parameters for Figure 4 are identical to those used for generating the curves in Figure 2, except that for these rats it was assumed that $P_1(A) = 1.0$: The stimuli were taken to be completely obvious, as they might be in a position problem. As before, the nonovertrained subjects were reversed after meeting a criterion of 15 consecutive successes, and the overtrained group was trained for 150 trials past this point and reversed. There is no ORE in Figure 4. The overtrained group starts reversal inferior to the nonovertrained one and holds this inferiority throughout reversal. Judging from the group learning curves, one would expect that overtraining might have produced a significant *decrement* in reversal speed for this obvious problem. It did. The nonovertrained group met the reversal criterion in 81.1 trials, as contrasted with 96.6 trials for the overtrained group. (A t test shows this difference to be significant, $p < .02$.) For this obvious problem, acquisition was much quicker than for the one previously discussed: Here the acquistion criterion was reached in 33.1 trials.

One more statistic of the stat-rat data is of interest: the number of initial errors before the first reversal success. For both problems the overtrained subjects made more such errors. For the animals with $P_1(A) = .5$, the means for overtrained and nonovertrained groups were 10.13 and 5.83, respectively. For the problem with $P_1(A) = 1.0$, these means were 11.12 and 6.0. Again, the stat-rat results closely parallel those found in experiments.

Discussion

The model which has just been described will, for reasonable parameter values, predict an ORE for stimuli which are fairly obvious but not completely obvious. The initial probability of attending, $P_1(A)$, can be varied from about .3 to about .7 without removing the ORE. For $P_1(A) < .3$, learning is slow, and a great many stat rats "refuse to learn." While it has been proven that such subjects would eventually learn, many hundreds of trials might be required before they did. With $P_1(A) = .4$, about 1 subject in 25 "fails to learn" in a reasonable length of time. As $P_1(A)$ increases past about .8, the difference between overtrained and nonovertrained groups disappears until with $P_1(A) = 1.0$ overtraining hurts reversal (see Figure 4). This gives us a clue about conditions when the ORE might not be found in the brightness situation: in apparatus where the stimuli are *very obvious*, so that the rat must almost always "notice" them, and $P_1(A)$ is very high. D'Amato and his group have repeatedly failed to find the ORE in his automatic Y mazes (D'Amato & Schiff, 1965). In this situation the whole positive arm is diffusely illuminated by the light cue. And the whole negative arm is 1,000 times darker. Thus the rats are literally "bathed" in the stimuli. Perhaps this discrimination is just too obvious for an ORE to appear.

There are other results from D'Amato's laboratory which indirectly confirm this interpretation. We have argued about the importance of the initial probability of attending to the relevant stimuli. In most T or Y mazes, a "position" problem involves motor cues (the actual turning response), abstract direction cues (going left or right), and extramaze cues (going to the *place* on the left, or the *place* on the right). This last class of cues (which would be most useful in forming a "cognitive map") is deliberately omitted from the automatic Y maze used in D'Amato's experiments. This might mean that "position" is a less dominant dimension in this apparatus than in most others (and hence that other cues, like brightness, are more available). Perhaps this is why in five pure position experiments, D'Amato's group has never found any significant decrement in reversal speed due to overtraining (1962, 1964). Clayton (1963a), Galanter and Bush (1959), Kendler and Kimm (1964), and Hill and Spear (1963) have found such a decrement. (So did Hill, Spear, & Clayton, 1962, in three separate experiments. But they reversed their animals for a short, constant number of trials, rather than to a criterion. Thus their results may have been influenced by the long string of initial errors which is generally made by the overtrained group in any overtraining reversal study.) In one experiment in the automatic Y maze, brightness was consistently correlated with position, making an especially obvious problem. In this case, and in this case alone, overtraining led to a significant decrement in reversal trials to criterion (D'Amato & Schiff, 1964). Perhaps this substantiates the argument that position, in D'Amato's apparatus, is a less dominant cue than in more standard apparatus.

Erlebacher also did a very careful brightness study and failed to find an ORE. If one looks at the various apparatus which have been used in brightness studies, one finds that Erlebacher's was the only enclosed T maze. It had stimuli painted on the sides, the ends, and the floors of the arms of the T. Thus in Erlebacher's study, too, the rat was surrounded by the relevant stimuli. Perhaps this problem, like D'Amato's, was too obvious. A great majority of the studies in which an ORE has been found have used stimuli painted on doors which the rat knocked down or opened in making his choice. Only one other experiment has used a T maze (Capaldi & Stevenson, 1957), and that was an elevated T, $1\frac{1}{2}$ inches wide. The rat presumably had other things to occupy his attention besides the color of the paint on the floor.

But if Erlebacher's and D'Amato's apparatus were such that $P_1(A)$ was high for brightness, then this fact should be reflected in other ways besides their failure to find an ORE. It is. We have argued that, in general, position is obvious and brightness is obscure. This difference affects the learning rates found in those two types of problems. Position is generally learned much more quickly than brightness. Hill and Spear, for example, found rats took about 17 trials to arrive at criterion in a T-maze position problem with a small food reward. (Contrast this with the 51 trials required for position learning in D'Amato's automatic Y maze—D'Amato & Jagoda, 1962, Experiment IV). Brightness, on the other hand, is usually learned to the same criterion in between 60 and 80 trials (Mackintosh, 1962; Pubols, 1956; Reid, 1953). D'Amato's rats fall into this range (D'Amato & Schiff, 1965), but Erlebacher's subjects reached criterion in 27.8 trials. Thus it appears that Erlebacher's animals learned brightness unusually quickly (indicating an easy, or obvious problem). D'Amato's animals learned position very slowly. This may indicate that position is an obscure dimension, and hence that brightness is relatively stronger; alternatively it may mean that learning is generally slow in that apparatus, in which case brightness is learned "too fast," and must have been obvious.

This post hoc theorizing does more than sweep the offending data under the rug: It also gives us a handle on the thorny problem of estimating $P_1(A)$ for a given situation. The obviousness of a problem should be reflected in its difficulty as compared with a standard position problem. If a problem is three or four times as difficult as a position problem, then one might well expect an ORE to appear. If it is only about twice as difficult, one could expect the ORE to be elusive. Experimentation across problems in a standard piece of apparatus is needed here.

In order to "wrap up" the existing data on the ORE, we assumed that position was a very obvious cue for a rat, so that in a position problem $P_1(A)=1$. We must now face the fact that Pubols (1956) found an ORE in a position problem as did Capaldi in 1963.

Both these studies were small; the first involved eight and the second, five rats in each group. Pubol's result seems to have been a statistical accident. His overtrained group learned the reversal more quickly than the nonovertrained group; but if one inspects his data, one discovers that the overtrained group also learned the initial problem about twice as quickly. One suspects that Pubols had smart rats in his overtrained group. Hill and Spear (1963) replicated the conditions of Pubol's study in detail and found no ORE, but rather a significant decrement in reversal trials to criterion because of overtraining.

Capaldi's study was smaller but is not so easy to explain away. Sutherland has pointed out[2] that on an elevated T maze like Capaldi's, rats are afraid, and rats under a strong fear drive will tend to fixate on one position in a T. Perhaps Capaldi's nonovertrained group, because it had had less experience on the maze, was more afraid and hence more fixated by fear. It should also be mentioned that Capaldi did not train his rats to a previously set performance criterion but trained all nonovertrained subjects for a constant 16 trials. Thus although Capaldi's study does not fit in with the other data considered above, the discrepancy may be due to somewhat special factors. Further research is needed to resolve this problem.

With minor exceptions, then, the present model seems to tolerate all existing data on the overlearning reversal effect.

MODEL II—AN ANALYZABLE MODEL

The model which has just been described seems promising. But it is unpleasant that mathematical analysis of Model I has not been possible. For this reason, a search has been made for a model which is similar to Model I in its assumptions but which is sufficiently different so that some analysis is possible. One such model has been found, and a complete report of the analysis appears elsewhere (Lovejoy, 1965). It was found that by making a seemingly innocent change in the assumptions of Model I it was possible to produce a model with many properties which can be determined analytically.

Model I stated that two interacting learning processes combine in discrimination learning: learning to attend, and learning to respond appro-

---

[2] N. S. Sutherland, personal communication, August 1964.

TABLE 3

TRANSFORMATIONS APPLIED AFTER EACH EVENT AND LIMIT VECTORS FOR THESE TRANSFORMATIONS

Probability event		$P(AW)$	$P(AB)$	$P(\tilde{A})$
AW	P_{n+1}:	$\alpha P_n(AW)+1-\alpha$	$\alpha P_n(AB)$	$\alpha P_n(\tilde{A})$
	P_∞:	1	0	0
AB	P_{n+1}:	$\alpha P_n(AW)+(1-\alpha)\lambda_1$	$\alpha P_n(AB)$	$\alpha P_n(\tilde{A})+(1-\alpha)(1-\lambda_1)$
	P_∞:	λ_1	0	$1-\lambda_1$
$\tilde{A}W$	P_{n+1}:	$\alpha P_n(AW)$	$\alpha P_n(AB)$	$\alpha P_n(\tilde{A})+1-\alpha$
	P_∞:	0	0	1
$\tilde{A}B$	P_{n+1}:	$\alpha P_n(AW)+(1-\alpha)\frac{1}{2}$	$\alpha P_n(AB)+(1-\alpha)\frac{1}{2}$	$\alpha P_n(\tilde{A})$
	P_∞:	$\frac{1}{2}$	$\frac{1}{2}$	0

priately after attending. Linear operators were introduced to describe probability changes at these two levels of learning. Model II is in a sense a single-level model: It assumes that on a trial an animal makes one of three responses. He attends and chooses white (Response AW), he attends and chooses black (Response AB), or he fails to attend (Response $\tilde{A}$). The attending response is totally bound to the two subsequent choices. In this simpler three-response model, any change in one probability will normally change both the others since the three must together sum to unity.

The state diagram for Model II is the same as that for Model I and is given in Figure 1. The same four events are possible: AW (rewarded), AB (nonrewarded), $\tilde{A}W$ (rewarded), and $\tilde{A}B$ (nonrewarded). The state of an animal in this three-response situation can be fully summarized by a three vector, $P_n=[P_n(AW), P_n(AB), P_n(\tilde{A})]$, where the three coordinates of the vector must sum to 1. For each of the four experimental events, a linear transformation of this three vector is assumed. These transformations and their limit values are given in Table 3. A full description of the justifications for these transformations is given in a separate paper (Lovejoy, 1965), but the basic ideas are: (*a*) When a response occurs and is rewarded, it becomes more probable, and the probabilities of both the other responses are reduced by a factor α $(0<\alpha<1)$; (*b*) when a response occurs and is not rewarded, its probability is decreased by the same factor, α, and the probability so lost is divided up among the other two responses in such a way that they approach constant limit points. Note, for example, that when Event $\tilde{A}B$ occurs (the animal fails to attend and is not rewarded), the probability of nonattending is reduced by the factor α, and the probabilities of the two other responses each approach a limit value of $\frac{1}{2}$. This value is the result of a symmetry assumption—since the subject was not attending to the relevant dimension, he cannot "know" whether he went to black or white, so he should treat the two colors symmetrically. When Event AB occurs, on the other hand, no such symmetry argument is applicable; accordingly the limit point must be left as a new free parameter in the model, λ_1.

More will be said below about the

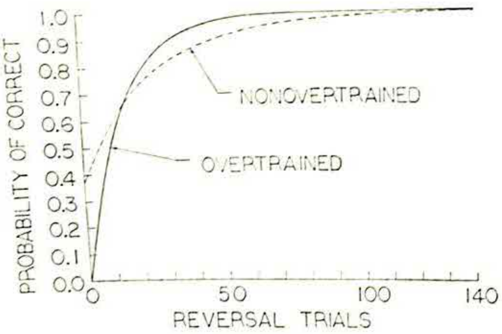

FIG. 5. Reversal learning curves, calculated according to Model II (Theorem 1) for two groups of animals with $P_1(A) = 0$.

mathematical and psychological importance of these assumptions and how they distinguish Model II from Model I. For the moment it is of interest to note that this model, too, can predict the ORE. Because the single-stage Model II is simpler mathematically, it has been possible to derive an explicit expression for the expected mean learning curve for this model. This is done in Theorem 1 of the companion paper. Using this theorem, it is a simple matter to investigate learning and reversal with various values of $P_1(A)$, the initial probability of attending. Since "attending" is not a unitary response in Model II, we must add $P_1(AW)$ and $P_1(AB)$ to find the analogue of $P_1(A)$ in Model I. Figure 5 is the Model II version of Figure 2 for Model I; the curves of Figure 5 are smooth because for this second model the exact form of the expected mean learning curve can be predicted, and so there is no need to resort to stat rats to explore the model. Figure 5 shows the course of reversal learning for two groups of Model II animals. The nonovertrained group was reversed after 20 trials of acquistion, the overtrained group after 300. These curves are for reversal of a very difficult problem, in that $P_1(A)$ was taken to be 0. The rate parameter, α, was taken to be .93 for these computations, and the limit point, λ_1, was taken to be 1. Under these conditions, the overtrained group clearly overtakes the nonovertrained group early in reversal. Like Model I, Model II predicts an ORE if $P_1(A)$ is low.

Figure 6 shows similar curves, calculated with the same learning parameters as Figure 5, but here the problem was very obvious: $P_1(A)$ was taken to be 1. The nonovertrained group was trained for 10 trials before reversal (learning is fast in this obvious problem); the overtrained group was reversed after 300 trials. For these subjects, overtraining has slowed the whole course of reversal learning. Again, Model II parallels the prediction of Model I.

The amounts of training given and the values chosen for $P_1(A)$ and λ_1 are rather extreme for Figures 5 and 6. These values have been chosen to accentuate the fact that this simple, single-process model can on appropriate occasions predict the ORE, and that at least qualitatively this model resembles the data which have been found. These two figures are representative of the types of curves

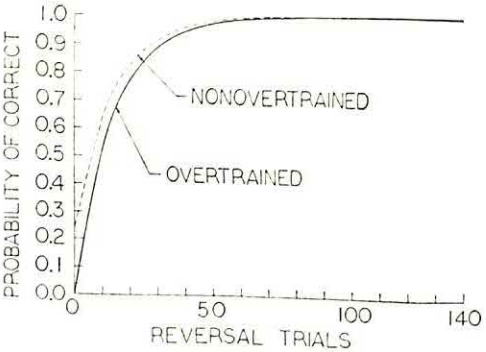

FIG. 6. Reversal learning curves, calculated according to Model II (Theorem 1) for two groups of animals with the same parameters as those in Figure 5, except that here $P_1(A) = 1.0$.

which are very generally found for a high value of λ_1: When $P_1(A)$ is high, the learning curves do not cross, but when $P_1(A)$ is low, they invariably do. Actual experimental conditions would probably correspond to less separated curves than those in Figures 5 and 6, but the same general shapes could be expected. Because of the tractability of Model II, it has been possible to *prove* the prediction of an ORE not only by means of curves like those shown in Figures 5 and 6, but also by investigating the relationship between the amount of prereversal training and the number of errors in reversal learning. These results, and also the predictions of Model II for extinction conditions, are derived and discussed in the companion paper.

Since Model II can be handled mathematically, the reader may well wonder why so much time has been given to a discussion of the less elegant Model I. This point is central in the following section.

A COMPARISON OF THE TWO MODELS

Two different models have been presented for discrimination learning, each of which is capable of predicting an ORE with some parameter values. For the first model it was assumed that "attending to the relevant dimension" was a response, which was strengthened by reward and weakened by nonreward. "Choosing white after attending" was also assumed to be a response and was assumed to change in a similar fashion. This model has considerable intuitive appeal, but it has been impossible to analyze mathematically. The second model assumed that the subject made only one response on any trial, which might be "attend and choose white," "attend and choose black," or "do not attend to the relevant stimuli." This change in definition of the response (or of the probability to be transformed) is the principal conceptual difference between the two models. But how does this difference affect the mathematical and psychological properties of the models?

A crucial assumption in Model I is that when the subject fails to attend, no change occurs in $P(W|A)$, his chance of correct choice on attending trials. This means that if for some reason a subject stops attending to the relevant stimulus for a number of trials, $P(W|A)$ will remain at a constant value, even though the animal may be responding in the experimental situation. $P(W|A)$ is dissociated from the learning process when the subject is nonattending. This dissociation is reminiscent of Krechevsky's notion of "hypothesis behavior," and has strong experimental implications which will be discussed below. It also makes mathematical analysis of Model I very difficult.

Model II, on the other hand, is closer to the "continuity" view of learning. All trials affect all probabilities. The crucial difference between the two models is the effect of Event $\tilde{A}B$: nonrewarded nonattention. In Model I this event makes the subject more likely to attend, since nonattending was nonrewarded, but it does change the relative strengths of the two attending responses. For Model II, too, Event $\tilde{A}B$ makes attending more likely. But at the same time it moves both $P(AB)$ and $P(AW)$ towards their limit value of $\frac{1}{2}$ (see Table 3). Consequently $P(W|A)$ moves towards a limit value of $\frac{1}{2}$, too. Thus, during a long string of trials when an animal is not attending to the relevant stimuli, Model II predicts that an animal "forgets" what he had previously learned about them. Model I predicts that the rat has "memory" in the form of an unchanged $P(W|A)$.

There is some evidence that a "dissociation" assumption like that in Model I may be necessary. Goodwin and Lawrence (1955), for example, ran rats in a T maze with two stimulus dimensions besides position (which was varied randomly). On each trial one arm was white and the other black; at the same time one arm had high hurdles, and the other low hurdles. These two cues were uncorrelated. In Stage I, animals learned a brightness discrimination. In Stage II, they learned to discriminate high from low hurdles but with brightness cues still present. In the third stage, they either relearned the Stage I brightness problem, or they learned the reversal. It was found that relearning was much faster than reversal, and that overtraining on Stage II did not change this result. It appears that the probability of going to white given that the subject was attending to brightness, $P(W|A)$, stayed high throughout the second stage, despite the fact that brightness cues were present and randomly rewarded during Stage II. Model I, with its dissociation assumption, could easily be extended to encompass such a result. It is not easy to see how to make a similar extension of the second model.

Another datum which adds to the plausibility of the dissociation assumption was first discovered by D'Amato and Jagoda (1960) and has since been confirmed by Mackintosh (1963c). These experiments, too, had three stages. In the first, rats learned a brightness problem to criterion. Then they were extinguished (neither response rewarded) either for a fixed number of trials (D'Amato & Jagoda) or to a criterion (Mackintosh). After this, they learned the reversal problem. Both experimenters found that just after reversal learning started, there was an abrupt drop in the proportion of correct responses (in the reversal). It appeared that these subjects had stopped heeding brightness in extinction, and that when they started to attend to brightness in reversal, they still had the habit of choosing the formerly correct alternative. It seems, again, that $P(W|A)$ had been dissociated during the extinction period.

A third reason for liking this dissociation assumption is that Model I seems to be more nearly a correct representation of the real world than Model II. The parameter values for the Model II curves shown were carefully chosen. Although the model often predicts an ORE, only in some rather special cases does it predict one which is as large as the experimentally obtained ORE. For Model I, on the other hand, it is easy to get an ORE. And the curves shown are honestly representative of the behavior of the model over a large par of the parameter space.

SOME FINAL DISCUSSION

In both Model I and Model II, the initial probability of attending, $P_1(A)$, corresponds to the "obviousness" or "distinctiveness" of the stimuli. This variable proved to be crucial in determining whether or not an ORE was predicted. And it also has the desired effect on speed of learning: The higher the initial probability of attending, the faster learning occurs. Indeed, one of the attractive features of the present models is that changes in the stimulus situation change the speed of learning by changing a perceptual parameter, $P_1(A)$, rather than by changing a learning-rate parameter —which might better depend on such variables as motivation, or the species of the subject. Results like those of Gibson and Walk (1956) can similarly be explained by assuming that "perceptual learning" affects $P_1(A)$. No change in rate parameter need be assumed.

Attending and Observing—Not the Same

A number of previous quantitative models for discrimination learning have been concerned with some sort of perceptual mechanism like the "attending" process in the present models. Wyckoff (1952) and Atkinson (1960) postulated overt observing responses which were affected by reward or nonreward much as was attending in this paper. Their formulations avoided unnecessary "mentalism," but there is now evidence that a more central process is involved.

Mackintosh (1962, 1964) has done a number of experiments which indicate that an *attending* response, rather than an overt observing response, is needed to account for the ORE. One of these was especially forceful. In 964, Mackintosh did a two-stage study. In the first stage, animals learned to distinguish a square from a diamond. The square was the positive stimulus, so it is clear that the rats learned all necessary overt responses for observing it. On half the trials both stimuli were black; on the other half, both were white. This change was irrelevant to the solution of the first problem, but ensured that the subjects were orienting towards the black and white squares. Half of the Stage I subjects were overtrained, and so, presumably, had learned the "observing" response especially well. Stage II was a brightness discrimination in which the stimuli were the same black and white squares which had appeared in Stage I. So if any overt observing or orienting response is crucial for the ORE, the overtrained subjects should have learned the new problem faster. In fact, they learned it more slowly. Mackintosh interprets this result as meaning that overtraining made the animal more likely to attend to spatial orientation—the relevant dimension for Stage I—and hence less likely to attend to brightness—the relevant dimension for Stage II. This interpretation agrees nicely with the present models.

Some Unsolved Problems

Before we can completely understand a rat's behavior in even the simple black-white discrimination learning situation, it appears that we must find satisfactory answers to a number of questions, some of which have been touched on above: How many dimensions can a rat attend to at any one time, or on any one trial? Both models have cunningly avoided this question by concentrating on attention or nonattention to a *single relevant dimension* and not concerning themselves with other dimensions. While this ploy has been moderately successful, one would hope eventually to expand his understanding. It has long been known that an animal can be responding consistently to one position and still be learning something about brightness (Spence, 1945). Indeed, while responding to position his latencies to black and white cards may indicate that he distinguishes them (Mackintosh, 1963a; Mahut, 1954; Maier, Glaser, & Klee, 1940; Ross, 1962). But if an animal thus appears to be attending to more than one dimension, is he learning about more than one? And how many dimensions interact in controlling behavior on any given trial? Perhaps only a single dimension is effective. In this case, how does an animal shift dimensions? If an animal is responding consistently along one dimension, does that necessarily mean that he is learning about that dimension? If more than one dimension is relevant, how do response strengths combine to shape behavior? What happens if two dimensions are opposed to each other? What if a dimension is only partially relevant (i.e., it helps the subject to do better than chance, but does not guide him reliably to reward)? Do animals always learn, or do they sometimes adopt a wait-and-see approach which may result in no change in state over a large block of trials?

These are all important questions, and we are certainly a long way away from solving most of them. It seems to the writer, however, that some sort of mathematical formulation will be most useful, and probably necessary, if an attack is to be made on many of them. As our questions get more subtle and we start to consider an increasing number of possibly relevant dimensions, Model II may be expected to come into its own. For Model I, attractive though it is, can only be appraised through stat-rat simulation. These techniques are possible, although not very powerful, in the simple case treated above. As more complicated situations are considered, and more parameters are introduced, mathematical tractability becomes more than a luxury; it becomes a necessity. It will always be possible to run a few groups of stat rats to demonstrate the plausibility of a given approach. But as soon as a large number of parameters is involved, mapping out the parameter space with stat-rat computations becomes prohibitively expensive. It seems certain that analytic techniques must be found. And by extending and exploring models like Model II, one dares hope they will.

REFERENCES

ATKINSON, R. C. A theory of stimulus discrimination learning. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences.* Stanford: Stanford Univer. Press, 1960. Pp. 221–241.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

CAPALDI, E. J. Overlearning reversal effect in a spatial discrimination task. *Perceptual and Motor Skills*, 1963, **16**, 335–336.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *Journal of Comparative and Physiological Psychology*, 1957, **50**, 195–198.

CLAYTON, K. N. Overlearning and reversal of a spatial discrimination by rats. *Perceptual and Motor Skills*, 1963, **17**, 83–85. (a)

CLAYTON, K. N. Reversal performance by rats following overlearning with and without irrelevant stimuli. *Journal of Experimental Psychology*, 1963, **66**, 255–259. (b)

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials on discrimination reversal. *Journal of Experimental Psychology*, 1960, **59**, 254–260.

D'AMATO, M. R., & JAGODA, H. Analysis of the role of overlearning in discrimination learning. *Journal of Experimental Psychology*, 1961, **61**, 45–50.

D'AMATO, M. R., & JAGODA, H. Overlearning and position reversal. *Journal of Experimental Psychology*, 1962, **64**, 117–122.

D'AMATO, M. R., & SCHIFF, D. Further studies of overlearning and position reversal learning. *Psychological Reports*, 1964, **14**, 380–382.

D'AMATO, M. R., & SCHIFF, D. Overlearning and brightness-discrimination reversal. *Journal of Experimental Psychology*, 1965, **69**, 375–381.

ERLEBACHER, A. Reversal learning in rats as a function of percentage of reinforcement and degree of learning. *Journal of Experimental Psychology*, 1963, **66**, 84–90.

GALANTER, E., & BUSH, R. R. Some T-maze experiments. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory.* Stanford: Stanford Univer. Press, 1959. Pp. 265–289.

GIBSON, E. J., & WALK, R. D. The effect of prolonged exposure to visual patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 239–242.

Goodwin, W. R., & Lawrence, D. H. The functional independence of two discrimination habits associated with a constant stimulus situation. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 437–443.

Harlow, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1949. Pp. 492–537.

Hill, W. F., & Spear, N. E. A replication of overlearning and reversal in a T maze. *Journal of Experimental Psychology*, 1963, **65**, 317.

Hill, W. F., Spear, N. E., & Clayton, K. N. T-maze reversal learning after several different overtraining procedures. *Journal of Experimental Psychology*, 1962, **64**, 533–540.

Kendler, H. H., & Kimm, U. Reinforcement and cue factors in reversal learning. *Psychonomic Science*, 1964, **1**, 309–310.

Komaki, J. The facilitative effect of overlearning in discrimination learning by white rats. *Psychologia*, 1961, **4**, 28–35.

Komaki, J. Reversal retardation through introducing forced nonrewarding trials in overtraining by white rats. *Annual of Animal Psychology*, 1962, **12**, 1–10.

Lovejoy, E. P. An attention theory of discrimination learning. *Journal of Mathematical Psychology*, 1965, **2**, 342–362.

Mackintosh, N. J. The effects of overtraining on a reversal and a nonreversal shift. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 555–559.

Mackintosh, N. J. Discrimination learning in animals. Unpublished doctoral dissertation, University of Oxford, 1963. (a)

Mackintosh, N. J. The effect of irrelevant cues on reversal learning in the rat. *British Journal of Psychology*, 1963, **54**, 127–134. (b)

Mackintosh, N. J. Extinction of a discrimination habit as a function of overtraining. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 842–847. (c)

Mackintosh, N. J. Overtraining and transfer within and between dimensions in the rat. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 250–256.

Mahut, H. The effect of stimulus position on visual discrimination by the rat. *Canadian Journal of Psychology*, 1954, **8**, 130–138.

Maier, N. R. F., Glaser, N. M., & Klee, J. B. Studies of abnormal behavior in the rat: III. The development of behavior fixations through frustration. *Journal of Experimental Psychology*, 1940, **26**, 521–546.

North, A. J., & Stimmel, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychological Reports*, 1960, 6, 227–234.

Pubols, B. H., Jr. The facilitation of visual and spatial discrimination reversal by overtraining. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 243–248.

Reid, L. S. The development of noncontinuity behavior through continuity learning. *Journal of Experimental Psychology*, 1953, **46**, 107–112.

Ross, L. E. The effect of equal reinforcement of the positive and negative discriminanda of a learned discrimination. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 260–266.

Siegel, S., & Wagner, A. R. Extended acquisition training and resistance to extinction. *Journal of Experimental Psychology*, 1963, 66, 308–310.

Spence, K. W. An experimental test of the continuity and noncontinuity theories of discrimination learning. *Journal of Experimental Psychology*, 1945, **35**, 253–266.

Sutherland, N. S. Stimulus analysing mechanisms. In, *Proceedings of a symposium on the mechanization of thought processes*. Vol. 2. London: Her Majesty's Stationary Office, 1959. Pp. 575–609.

Wyckoff, L. B., Jr. The role of observing responses in discrimination learning: Part I. *Psychological Review*, 1952, **59**, 431–442.

Wyckoff, L. B., Jr. A mathematical model and an electronic model for learning. *Psychological Review*, 1954, **61**, 89–97.

Zeaman, D., & House, B. J. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook in mental deficiency: Psychological theory and research*. New York: McGraw-Hill, 1963. Pp. 159–223.

(Received October 9, 1964)

Psychological Review
1966, Vol. 73, No. 1, 104–110

THEORETICAL NOTES

PIAGET ON PLAY:
A CRITIQUE

BRIAN SUTTON-SMITH[1]
Bowling Green State University

It is contended that Piaget's attempt to reduce play to a function of thought is vitiated by his concern with a restricted class of intellectual operations and by his copyist epistemology. By resting the origin of representation on a copyist notion of imitation and by permitting play only the function of replicating concepts, Piaget deprives play of any genuinely constitutive role within thought. The contradictory consequences both in terms of the evidence about play's ontogenesis and Piaget's ultimate resort to more characteristic affective and conative explanations are discussed.

Piaget's (1951) *Play, Dreams and Imitation in Childhood* is the most recent of many attempts to make sense out of the theoretically anomalous subject matter of play. With the possible exception of psychoanalytic play theory, it is the most conceptually elaborate account of play yet to be presented and, in addition, includes the best available examples of the sequence of play activities in the first years of life. It is the intention of this paper to argue that Piaget's thesis, to the effect that play may be interpreted functionally as an activity subordinated to adaptive intelligence, leads to contradictions within his own system. The present analysis will maintain that Piaget's epistemological assumptions and his concern with directed thought lie at the base of his misconstrual of play.

PIAGET'S EPISTEMOLOGICAL ASSUMPTIONS

Piaget's treatment of play is inseparable from his theorizing on intelligence. His root metaphor for the origins of intelligence in human beings is the process of biological adaptation in lower organisms. He is said to have been impressed in his earlier studies of mollusks by the way in which these lower organisms, while accommodating to the environment, also actively assimilated it in accord with their own schemas of action.

> An image of an active organism which both selects and incorporates stimuli in a manner determined by its structure, while at the same time adapting its structure to the stimuli, emerged from these early studies as a ready made model for cognitive development [Flavell, 1963, p. 36].

The concepts of assimilation and accommodation which were said to be applicable to lower organisms were also claimed to have applicability to the "intelligent" activities of higher organisms. These two became the indissociable functional invariants in Piaget's conceptual system for the description of intelligence:

> Accommodation of mental structure to reality implies the existence of assimilatory schemata apart from which any structures would be impossible. Inversely, the formation of schemata through assimilation entails the utilization of external realities to which the former must accommodate . . . [1954, pp. 352–353].

---

[1] This paper owes much to the stimulation of a visiting associate professorship in the Psychology Department and the Heinz Werner Institute of Developmental Psychology at Clark University, 1963–64. I am particularly indebted to Joseph Glick of Yale University and Bernard Kaplan of Clark University for their comments and suggestions.

Having made assimilation and accommodation ultimate categories of adaptive intelligence, Piaget sought to express other (related) functions in these same terms. Imitation and play came to be formulated as particular types of relationships between accommodation and assimilation. When accommodation to external reality dominated over assimilation, there existed the state which Piaget termed imitation. Alternatively, if assimilation of external reality to preexisting concepts occurred, then there was the state to which the term play was given. Imitation and play were said to be polarized examples of the activity of intelligence: "Intelligent adaptation, imitation and play are thus three possibilities and they result according as there is a stable equilibrium between assimilation and accommodation or primacy of one of these two tendencies over the other [1951, p. 86]." From this description it follows that imitation and play should be explicable in terms of the same processes that determine adaptive or intellective functioning. Leaving aside whether Piaget has shown that imitation is explicable in terms of his categories of adaptive functioning, it will be argued that his attempt to treat play within these categories is inadequate.

THE ASYMMETRY OF IMITATION AND PLAY

There are Hegelian overtones in Piaget's view that imitation and play are dialectical opposites which are gradually reconciled in a higher synthesis as the adaptive intellectual structures of logic progressively succeed them. But closer examination of the respective roles of these two functions within his system suggests that imitation and play are not related as equipotent thesis and antithesis. Piaget verges, it is true, on the suggestion that in the origin of representative thought, imitation supplies the images and play supplies the symbols. But this antiphonal relationship is alluded to ambiguously [1951, p. 3; 1960, p. 126], and, for the greater part, only imitation is given a key role in the origin of concepts. Imitation is said to be a source of representation, but play is used largely as an *illustration* of what representation is like.

Piaget has criticized some of his predecessors, among other things, for their copy theory of knowledge or their naïve realism. He has insisted that representative thinking is not merely a copy of external reality, but is rather the outcome of a long process of interaction between the organism and the environment with each modifying the other over time. As indicated above, he has stated this interaction in terms of the concepts of assimilation and accommodation. It is very clear from the corpus of his writings that he does not favor a copyist epistemology. Paradoxically enough, however, when the relationships between imitation and play are examined closely, the view emerges that Piaget has implicitly imported just such a copy theory into his own explanation of the processes of imitation, and that it is this implicit copyist notion of imitation which ultimately leaves play intellectually functionless within his system. To illustrate: The process of accommodation is portrayed as establishing for the subject a photographic-like negative of external reality. Imitation is said to be an extension of such negatives into positive action. "The function of imitation seems to be to produce this set of 'positives,' which correspond to the 'negatives' of accommodation and which, at each new 'printing,' make new reconstitutions and anticipations possible [1951, p. 84]." In due course, it is said, these imitations become interiorized as images, and then, later, when these images become attached to, and differentiated from, external symbols during intelligent activity or during play, they are transformed into concepts. It follows that sustaining and initiating this whole process of representative activity are the encapsulated photographic negatives which occur during accommodation. It may be true to say, as Piaget does, that this accommodation is "an *active* copy and not a trace or a sensory residue of perceived objects [1960, p. 126; italics mine]," but accommodation is nonetheless conceived as a *copy* and given the initiating and sustaining role in his formulation of representative activity.

Now it is conceivable that Piaget intends to say that, when the interiorized images become attached to symbols in adaptive intelligent activity, this process of attachment changes the interiorized images. But whether or not this is the case, and this is not relevant here, it is certainly true that in Piaget's systematic accounts of play such an attachment of symbols is not said to bring about any such change in images. Piaget's definition of play as predomination of assimilation over accommodation specifically precludes any such change. Instead the encapsulated images derived from accommodation determine the character of the activity. Play merely diversifies the symbols. As "reproductive assimilation," or "assimilation of reality to the ego," play can merely repeat; it can never originate. We thus have a situation in which the symbols of play are merely the reproductions of images preestablished through the copyist activity of imitation following accommodation. On this interpretation then, imitation is an essential factor in the constitution of representative activity, whereas play is not. It has no essential role in the structure of intellect as conceived by Piaget. Intelligence cannot proceed without imitation. It can proceed without play. It is hardly necessary to emphasize that there are alternative points of view which grant to the processes of symbol formation a more active role in the constitution of thought than this copyist theory of Piaget's seems to imply (Werner & Kaplan, 1963). Greater internal cogency is given to the present critique, however, by showing that this implicit copyist notion of the origin of concepts leads to peculiarities within the system itself which are neither self-consistent nor consistent with the known data concerning play.

THE DISJUNCTION OF COGNITIVE AND AFFECTIVE FUNCTIONS

Having deprived play of an active, intellective role in the constitution of new concepts, Piaget is forced into a series of substitute explanations. Unlike imitation which has the one definition, play is provided with a variety of definitions. At times it is referred to as "reproductive assimilation" (1951, p. 87), as "generalizing assimilation" (1951, p. 100), but also as "pure assimilation" (1951, p. 87), "mere assimilation" (1951, p. 89), and "distorting assimilation" (1951, p. 102). The latter three terms do not occupy any systematic status within Piaget's theory of intelligence as do the first two. They are never defined, and it is not entirely clear just what they can mean. It is this lack of a functional role for play within the structure of intelligence which apparently forces Piaget to find alternative nonintellective functional explanations for play, which he does not have to provide for imitation. It is said, for example, that play exists in childhood because thought is not adequate to its tasks:

> The prevalence of play among children is therefore to be explained not by specific causes peculiar to the realm of play, but by the fact that the characteristics of all behaviors and all thought are less in equilibrium in the early stages of mental development than in the adult stage [1951, p. 147].

> But why is there assimilation of reality to the ego instead of immediate assimilation of the universe to experience and logical thought? It is simply because in early childhood this thought has not yet been constructed, and during its development it is inadequate to supply the needs of daily life [1951, p. 166].

These quotations make it clear that play is not considered as making any intrinsic contribution to thought, but is instead some sort of compensation for thought's inadequacy. This interpretation is made explicit when Piaget says that the function of play is to serve ego continuity. "It follows that for the child assimilation of reality to the ego is a vital condition for continuity and development precisely because of the lack of equilibrium of this thought, and symbolic play satisfies this condition both as regard to signifier and signified [1951, p. 166]." But, surely, it is difficult to maintain that play is merely a polarity of thought, that the "prevalence of play . . . is therefore to be explained not by . . . causes peculiar to the realm of play," and then to invoke this concept of ego continuity which is not a concept intrinsic to Piaget's account of the nature of intelligence. This invocation of "ego continuity" does, in fact, suggest that play has some other function to serve which is quite distinct from thought. The notion of ego continuity as the function of play bears close resemblance to the concept of ego mastery as presented in psychoanalytic accounts, particularly that of Erikson (1950). But to rest play's positive contribution on such grounds is to lead to a disjunction of cognitive and affective functions, quite inconsistent with Piaget's own repeated stress that the two are not separable. If play has a constitutive affective function of this sort (ensuring ego continuity), Piaget's general position on affective-cognitive relations requires that it should also have a constitutive intellective function.

THE DISJUNCTION OF DIRECTED AND UNDIRECTED THINKING

It would be wrong to imply that Piaget's particular treatment of play stems perhaps from an accident in his epistemology. The approach is more deep-seated than that for there is a consistency between his account of infant thought and his account of adult thought. It has been shown that in his treatment of infant thought he gives an initiating role to images which derive from accommodation to external reality. In adult thought also he is concerned with those coherent formal structures of intellect that have to do with accommodation to and control over the external world. In particular he is concerned with the correspondence between "the structures described by logic and the actual thought processes studied by psychology [1957, p. xvii]." He is not concerned with those less directed aspects of adult thought usually referred to by such terms as reverie, creative imagination, or divergent thinking. Yet in the opinion of some, the latter have a great deal to do with novel forms of adaptation. It would not be farfetched to speculate, in fact, that if there is an intrinsic relationship between play and thought, it is more likely to be with these latter forms of divergent intellectual operations than with the directed forms which concern Piaget.

Piaget's formulation of adult thought has a number of consequences which are consistent with the epistemological orientation already outlined. For example, his view that the thinking of the young child is relatively disequilibrial is again similar to the empirical view which says that since thought copies reality, young children who are inexperienced will have more imperfect copies than older persons. Of course, stated in Piaget's terms, the young organism plays a more active role in acquiring its copies than was the case in the traditional empirical view, and unlike that view there is a telic quality to Piaget's formulation of the young child's operations. They appear to be "successive approximations to a kind of ideal equilibrium or end state never completely achieved [Flavell, 1963, p. 47]." The difficulty with this approach for Piaget's own point of view is that it is not consistent with his stage theory which says that each stage of thinking has a characteristic organization and completeness of its own. It is difficult to see how there can be both "equilibration" of early stages and the view that these early stages are relatively less in equilibrium than later stages. The problem can be located in the different usages of the term "equilibrium" which Piaget employs both to refer to characteristics of intellectual organization at one time or stage and to refer to the ontogenetic status of the organism in general. Apart from the semantic confusion in having such entities as "disequilibrial equilibriums," either stage or ontogenetic usage leads to difficulties with respect to the definition of play as a polarity of thought.

If it is said that when an intellectual structure is disrupted by intractable data the lack of intellectual resources in childhood (the disequilibrial ontogenetic state) leads to the use of play as an affective holding action, this is again to give play no genuine cognitive function. Yet, if it said that play is a true polarity of thought in the sense of reestablishing equilibrium through genuinely cognitive means at any one stage (the implied stage theory), this is equally unfortunate for the theory as it stands. If, for example, some divergent thinking function is attributed to the "distorting assimilations" of play, this requires a reconsideration of the role of divergent thinking not only in children's thinking but also in the development of adult thinking. If play does not simply "distort" old concepts, whatever such a phrase can mean, but also originates new concepts, then the whole account of the genesis of intellectual structures from childhood through adulthood has to be changed.

Piaget appears to have chosen rather to make a disjunction between the operations of thought in adulthood and the operations of thought in childhood, in addition to the formal differences between sensory-motor, concrete, and abstract operations which are already a part of his theory. Thus it is said that childhood thinking is bolstered by the affective operations of play, but the thinking of adults is not. From Piaget's viewpoint as undirected thinking, fantasy, play, etc., are specifically childlike and mainly compensatory, that is, having nothing to do with the development of particular kinds of intellectual operations, they may be confined to the infantile stage and regarded as irrelevant to the nature of adult intellectual operations.

THE ONTOGENESIS OF PLAY

It is legitimate to ask whether the evidence on play supports Piaget's point of view. His view that play is merely a buttress to an inadequate intelligence leads to the corollary that as that intelligence increases in efficiency and adequacy, play will cease to be important in the development of the mind. In Piaget's unilinear account of the ontogenesis of intellectual structures, play has such a transient position. Thus it is said that from about the age of 4 years onwards symbolic games decline because the child becomes capable of greater adaptive activity and is more realistic: ". . . progress in socialization instead of leading to an increase in symbolism transforms it more or less rapidly into an objective imitation of reality [1951, p. 139]." "Rule games . . . mark the decline of children's games and the transition to adult play, which ceases to be a vital function of the mind when the individual is socialized [1951, p. 168]." On the basis of the evidence about play several objections can be raised to the view that the symbolic games of childhood are simply replaced by more realism with age, and that play which is "vital" in childhood ceases to be vital in adulthood. In general it can be maintained with equal cogency that, rather than a decrease in the symbolic play function with age, what we actually find is a shift in the applications and the differentiation of this function. First, the early rule games of children continue to be heavily loaded with symbolic elements which play a part in determining outcomes equal in importance to skill. In the host of singing games of young girls, for example, it is the one who is arbitrarily chosen as Punchinello or as Farmer-in-the-Dell who determines the outcome. Similarly, in the great group of central-person games of early childhood which model the family drama, it is the player chosen by the arbitrary counting-out processes or the player with the role vestments of Caesar or Red Rover who decides by fiat when the game will start, when the players will run, etc. (Sutton-Smith, 1959). Here the play symbolism has been collectivized, but it has hardly been decreased, or given way to greater realism. In Piaget's (1932) treatment of games he was concerned with a game (marbles) in which the symbolism was much less pronounced, marbles being more age appropriate at the end of early childhood, around 9–10 years, whereas these other games just mentioned tend to flourish from 7 through 9 years. Furthermore, in that treatise, play at games was used to illustrate moral and social development rather than to provide the basis for an understanding of the play function. In a sense, Piaget (1932) first uses games to illustrate the development of morality, while he later (1951) uses play to illustrate the development of thought. Neither time does he deal systematically with the peculiar functions of games or play in human development. Perhaps it is this tendentiousness of approach that leads Piaget to discount the continuing influence of play in psychological development. Even when children's games are mainly matters of physical or intellectual skill, as from 10 years onwards, however, it is difficult to accept Piaget's view that this is simply a matter of play being converted from a symbolic to a realistic application. The total game performance still remains basically a symbolic procedure. There is, after all, little that is precisely rational about a game of football or a game of basketball. In fact, there is support for the point of view that such sports are best conceived as modern ritualized dramas of success and failure (Sutton-Smith, Roberts, & Kozelka, 1963) and as such have little that is either rational or real about them. Furthermore, it is the thesis of Huizinga (1949) that play (if not sport) does continue to be a vital function of mind even in socialized adults and that it permeates all their vital activities including those of a cultural sort and those of an economic and political nature. Recent cross-cultural work showing the full implication of games in culture would appear to substantiate this view (Roberts, Arth, & Bush, 1959; Roberts & Sutton-Smith, 1962; Sutton-Smith & Roberts, 1964). Though preschool egocentric symbolic play may decrease with age, play nevertheless finds expression in the midst of a variety of other cultural and social forms. It is thus not displaced by realism or by greater rationality, nor does it cease to be a vital function with age. Instead it becomes more differentiated and more representative in its contents of the other forms of human development. Without such a point of view it is difficult to understand the verbal play of adults, their social and sexual play, their rituals and their carnivals, their festivals and fairs, and their widespread and diversified playfulness.

This present criticism of Piaget's view of play as a transient, infantile stage in the emergence of thought parallels Vygotsky's (1962) criticism of Piaget's notion of egocentric speech as a transient, infantile stage in the emergence of socialized speech. Vygotsky maintained that in the course of ontogeny egocentric speech does not give way to socialized speech, but becomes transformed into inner speech which may itself be differentiated according to autistic or logical ends. Similarly, it has been maintained here that children's play does not give way to intelligent adaptation, but becomes differentiated in a variety of ways. The difference between Vygotsky's criticism and the present one, however, hinges on the fact that both Piaget and Vygotsky are talking about the ontogenesis of the same function—speech—whereas here two functions are involved, intelligence and play, which Piaget attempts to amalgamate into a single function (intelligence) and an aborted variant (play). It should be noted in conclusion that Piaget's various earlier treatments of undirected thinking in children, their animism, artificialism, etc., as transient, infantile stages in development are susceptible to similar criticism (Piaget, 1928, 1929, 1930).

CONCLUSION

It has been contended that Piaget's major concern with directed rather than undirected intellectual operations together with his implicit copyist epistemology have made it impossible for him to deal consistently with play. He has attempted to make it a function of thought without giving it any intellective function within thought. And he has been unable to give it such a function within thought because in his epistemology concepts are ultimately copies derived from and appropriate to an external reality. As play does not copy but "distorts" reality, it can have no intrinsic place in such an epistemology. As a result, Piaget is forced finally to explain play in terms of the type of affective and conative functions that are familiar in other play theories. (This is not to imply that there are not parallels between cognitive operations and play contents at given ages. It is the virtue of Piaget's observations that in this respect they complement those of psychoanalysts and sociologists which tend to be highly selective in terms of play content reflecting their own characteristic theoretical orientations. But the existence of such parallels between the operations of thought and the structures of play conveys nothing in itself about the function of play.)

While it has not been the purpose of this paper to propose any alternative theoretical approach to play or to criticize Piaget's play theory from alternative points of view, a brief mention of an alternative epistemology and the quite different approach to play which it entails may serve to highlight the legitimacy of the present criticism in epistemological terms. Ernst Cassirer (1953) explicitly opposes the view that concepts are copies of external reality. They are rather modes through which reality is constructed on a level transcending the simple sign functioning of lower organisms. From an epistemologically constructivist viewpoint of this sort play might well be taken as a very positive illustration of the thinking process because it involves the construction of symbols to create new conceptual domains. Though from Cassirer's point of view, play would have to be just one of many such symbolic modes with its own characteristic properties, others being art, myth, science, etc., for Cassirer's point of view with respect to symbolic development is not only constructionist, it is also multilinear. Piaget with his contrasting copyist assumption and his unilinear notion of thought is forced to reduce all forms of symbolic activity to the single copyist form of adaptive thought. And on the evidence of the internal inconsistencies to which this leads, his approach does not appear to have been fruitful for the explanation of play.

The present criticism of Piaget's theoretical system which focuses on the inadequacy of his epistemology when applied to play probably has relevance for the interpretation of divergent-thinking activities in general, such as creativity, originality, expressiveness, etc. At the same time this criticism probably has less relevance to such convergent-thinking operations as the understanding of physical causality, spatial relations, etc., with which Piaget is mainly concerned and for which his copyist assumption may be more intrinsically relevant. For whatever may be said here about Piaget's theory of play, there is no intention to detract from the conceptual flexibility of his system when applied to the child's understanding of the operations of the physical world. In that area man and mollusk share a common concern, and Piaget's epistemology appears to be revealing.

REFERENCES

Cassirer, E. *The philosophy of symbolic forms.* Vol. 1. New Haven: Yale Univer. Press, 1953.

Erikson, E. H. *Childhood and society.* New York: Norton, 1950.

Flavell, J. H. *The developmental psychology of Jean Piaget.* New York: Van Nostrand, 1963.

Huizinga, J. *Homo ludens: The play element in culture.* London: Routledge & Kegan Paul, 1949.

Piaget, J. *The language and thought of the child.* New York: Harcourt, Brace, 1926.

Piaget, J. *Judgment and reasoning in the child.* New York: Harcourt, Brace, 1928.

Piaget, J. *The child's conception of the world.* New York: Harcourt, Brace, 1929.

Piaget, J. *The child's conception of physical causality.* London: Kegan Paul, 1930.

Piaget, J. *The moral judgment of the child.* London: Kegan Paul, 1932.

Piaget, J. *Play, dreams and imitation in childhood.* London: Heinemann, 1951.

Piaget, J. *The construction of reality in the child.* New York: Basic Books, 1954.

Piaget, J. *Logic and psychology.* New York: Basic Books, 1957.

Piaget, J. *The psychology of intelligence.* Paterson, N. J.: Littlefield & Adams, 1960.

Roberts, J. M., Arth, M. J., & Bush, R. R. Games in culture. *American Anthropologist,* 1959, **61**, 597–605.

Roberts, J. M., & Sutton-Smith, B. Child training and game involvement. *Ethnology,* 1962, **1**, 166–185.

Sutton-Smith, B. *The games of New Zealand children.* Berkeley: Univer. California Press, 1959.

Sutton-Smith, B., & Roberts, J. M. Rubrics of competitive behavior. *Journal of Genetic Psychology,* 1964, **105**, 13–37.

Sutton-Smith, B., Roberts, J. M., & Kozelka, R. M. Game involvement in adults. *Journal of Social Psychology,* 1963, **60**, 15–30.

Vygotsky, L. S. *Thought and language.* Cambridge: Massachusetts Institute of Technology Press, 1962.

Werner, H., & Kaplan, B. *Symbol formation.* New York: Wiley, 1963.

(Received August 31, 1964)

Psychological Review
1966, Vol. 73, No. 1, 111–112

RESPONSE TO BRIAN SUTTON-SMITH

JEAN PIAGET

Faculté des Sciences, Université de Genève

The interesting criticism that Sutton-Smith has made of my play theory calls, I feel, for a brief reply. I must confess that I have some difficulty in recognizing the opinions that this author attributes to me, and I think that his formulations of them derive from the fact that he has only been able to assess that portion of my work that has been translated into English.

I have never, either implicitly or explicitly, expressed the view that knowledge is a copy of reality because such a view is contrary to my position with respect to the nature of intelligence. My whole conception of intellectual operations is based on the premise that to know or to understand is to transform reality and to assimilate it to schemes of transformation. In particular, I have never said that representative or symbolic thought, including concept formation, is derived from imitation (in the form of an accommodative copy). On the contrary, in my view, all concepts are derived first from the action and then from the operation, which is another way of saying that concepts are the expression of an assimilation by schemes of transformation.

In fact, the reality that intelligence tries to grasp consists of a series of states (A,B,C, etc.) and transformations which modify these states (A into B, B into C, etc.). One can, therefore, distinguish two components of cognitive functions:

There is first of all a figural component that does not itself constitute a copy but rather a more or less approximate description of reality states and their configurations. This figurative component is derived from perception, imitation, and imagery (graphic or mental) or from interiorized imitations.

There is secondly a cognitive component which takes account of transformations and which builds upon sensorimotor actions, interiorized actions, and finally, thought operations which are derived from actions and not at all from imitation.

In other words, when studying perception, mental imagery, etc. (research not yet published in English), I have tried to show that the figural aspects of cognitive functions are never sufficient to explain representative or conceptual knowledge. Indeed, the figurative components only play a useful role to the extent that they are subordinated to the cognitive component. It seems clear to me that knowledge of a "state" cannot be reduced to a copy of this state, but necessarily consists in an assimilation of this state to a preoperational scheme or a postoperational concept. In fact, to understand a state one must understand the transformations from which the state results which, let us repeat, excludes all types of copy knowledge.

The role of play in this system seems clear. Play is an exercise of action schemes and therefore part of the cognitive component of conception. At the same time, however, play manifests the peculiarity of a primacy of assimilation over accommodation which permits it to transform reality in its own manner without submitting that transformation to the criterion of objective fact. Regardless of what Sutton-Smith says, play fits into this system without becoming subordinated to accomodative imitation. Imitation only plays the role of a symbolic instrument from the moment that sensorimotor play becomes symbolic. This last point leads us to an examination of the symbolic function.

The sensorimotor functions (including sensorimotor play and perception) are not symbolic inasmuch as they only use undifferentiated signifiers (indexes or signals). The symbolic function begins around $1\frac{1}{2}$–2 years of age with the appearance of differentiated signifiers (signs and symbols) and consists in representing various figurative and cognitive schemes by means of these signifiers. In other words, the symbolic signifiers are derived from imitation which, before becoming interiorized, is already a kind of symbolization in action. This fact, however, in no way implies that symbolic instruments should be confused with figural aspects of thought. Perception is figurative, but not symbolic, while language is symbolic (in the broad sense), but not figurative. Interiorized imitation and images are, on the contrary, at the same time figurative and symbolic. Mental imagery in particular is the product of interiorized imitation and not the simple residue of perception as it was previously believed.

If symbolic play uses imitation, it is exclusively as a symbolic instrument. This follows because there are only two ways that an absent situation can be represented; it can either be described by language or evoked by imitative gestures or images. This in no way means, however, that symbolic play can be reduced to imitation since play is exclusively an assimilation of reality to the self. Nonetheless, since it is symbolic it needs signifiers, and it borrows them either from language or from the only other source of symbols, that is to say, gestural or interiorized imitation.

It seems clear now that the criticisms that Sutton-Smith has made of our theory are unfortunately the result of a series of misunderstandings and that the opinions that he attributed to us, which indeed deserve to be criticized, are fortunately not our views. There remains, however, one other point that must be clarified. The author defends the idea that symbolic play does not diminish but rather differentiates during the course of development. This is certainly true in one sense of the word. In becoming differentiated, however, it at the same time becomes more and more adequately adapted to reality (construction games, etc.). This is the sense in which I speak of play as diminishing with age. And it does diminish if one takes, as I do, the essential property of play to be the deformation and subordination of reality to the desires of the self.

Finally, Sutton-Smith makes an allusion to Vygotsky's conception of egocentric language. I have already replied in length to Vygotsky's insightful remarks. However, if we are to take up the problem of egocentrism again, which goes hand in hand with the deforming processes of assimilation, we must also situate language and play in the overall context of the individual and social actions of the child. In studying the collaboration of children in collective actions, R. F. Nielsen has found an evolutionary law which resembles to a great extent what I used to call the passage from egocentrism to cooperation.

(Received February, 1965)

VOL. 73, NO. 2 MARCH 1966

PSYCHOLOGICAL REVIEW

DRIVE THEORIES AND STIMULUS GENERALIZATION[1]

LOWELL H. STORMS
Neuropsychiatric Institute, University of California at Los Angeles Center for the Health Sciences

AND

WILLIAM E. BROEN, JR.
University of California, Los Angeles

Careful study of the implications of Estes' theory and D × H theories relating drive to stimulus generalization reveals that highly varied predictions result from changing the specification of boundary conditions. Consistent predictions from Estes' theory are flatter gradients of stimulus generalization (GSGs) with increased drive and steeper GSGs with decreased drive. However, the D × H theories predict either flatter or steeper GSGs with increased or decreased drive, and all theories predict that either raised or lowered GSGs will result from increased or decreased drive, depending upon boundary conditions. The boundary conditions and relevant variables which must be considered in making rigorous deductions are described. These include the relative predominance of stimulus sampling probability or energizing effects of drive as compared with drive stimulus change effects, initial and final levels of drive, strengths of competing responses, and height of the threshold. Since there are difficulties in measuring these variables, clear predictions from the theories will often be impossible.

Relationships between drive and stimulus generalization have been central considerations in theoretical formulations and investigations dealing with a variety of areas, such as approach and avoidance behavior (Brown, 1942, 1948), generalization of operant responses (Jenkins, Pascal, & Walker, 1958; Thomas & King, 1959), displacement (Miller & Murray, 1952), and schizophrenic behavior (Broen, Storms, & Goldberg, 1963; Mednick, 1958; Rodnick & Garmezy, 1957). In spite of the considerable theoretical importance and experimental focus on relationships between drive and gradients of stimulus generalization (GSGs), conclusions regarding these relationships have been varied and disparate. Increased drive has been credited with such effects as increasing elevations of GSGs (Mednick & Freedman, 1960), increasing (sharpening) their slopes (Spence, 1958), flattening GSGs (Estes, 1958; Jenkins et al., 1958), and sharpening and flattening GSGs as well as raising and lowering them (Thomas & King, 1959). Estes' stimulus sampling theory and $D \times H$ energizing theories such as those of Hull and Spence yield different predictions regarding the effects of drive changes on GSGs, partially depending upon conditions other than drive conditions (Estes, 1958, 1959b; Hull, 1943, 1952; Spence, 1956, 1958). It is not at all clear that the apparent inconsistencies in the evidence are inconsistent with any of these theories since quite different results might be predicted for differing experimental conditions. It is the purpose of this article to make explicit the implications of these theories for relationships between drive and stimulus generalization and to discuss the variables and conditions relevant to designing appropriate experimental tests.

---

[1] Preparation of this article was partially supported by Research Grant R 61-2-20 from the State of California Department of Mental Hygiene. The authors would like to express appreciation to Norman Anderson for his critical reading of an earlier draft of the article and many helpful comments and to Irwin Levin for computational aid.

Our interest will focus on the effects of drive on GSGs where a response has been trained to a stimulus or class of stimuli, and tests are made with the training stimulus and other stimuli on the stimulus continuum under various drive conditions. Thus, the effects of drive on discrimination learning and tests involving simultaneous presentation of more than one stimulus will not be considered. Although one response is trained to a single stimulus situation, all theories under consideration acknowledge that other responses may possess some probability of occurrence, or some strength (except, according to Estes, when some response has a unit probability of occurrence). Therefore, each stimulus will be considered associated with more than one possible response.

ESTES' THEORY

In Estes' (1959b) theory, on any trial only a sample of the set of stimulus (*S*) elements of which a stimulus is composed act on the receptors of an organism. Whenever a given *S* element is sampled just prior to a reinforced response, that *S* element becomes conditioned to the response. Conditioning is at full strength after one such trial. At any time, an *S* element can be conditioned to no more than one response. *S* elements available for sampling are sampled randomly according to their sampling probabilities. The probability of a response is equal to the proportion of the available *S* elements which are conditioned to the response, with all elements weighted by their sampling probabilities. Response probability can be increased by adding conditioning trials so that more *S* elements become conditioned to the response, by increasing the proportion of available *S* elements conditioned to the response (e.g., bringing in an additional stimulus previously established as a conditioned stimulus for the response), or by increasing the sampling probabilities of the *S* elements. Response probability can be decreased by decreasing the proportion of available *S* elements conditioned to the response, or decreasing the sampling probabilities of some or all of the *S* elements conditioned to the response.

Drive Effects in Estes' Theory

Although Estes (1959b) does not discuss drive in his formulations of his general theory, it is well established that drive changes have stimulus effects. Thus Estes (1958) offers a separate, straightforward interpretation of drive effects in terms of his stimulus sampling theory. Basically, drive changes result in altered response probabilities because of changes in the drive stimuli available for sampling and changes in the sampling probabilities of drive stimuli.

Intended to apply only to hunger and thirst, Estes' formulation of drive effects involves four classes of cues: satiation cues (*F*), such as stomach distension; deprivation cues (*D*), such

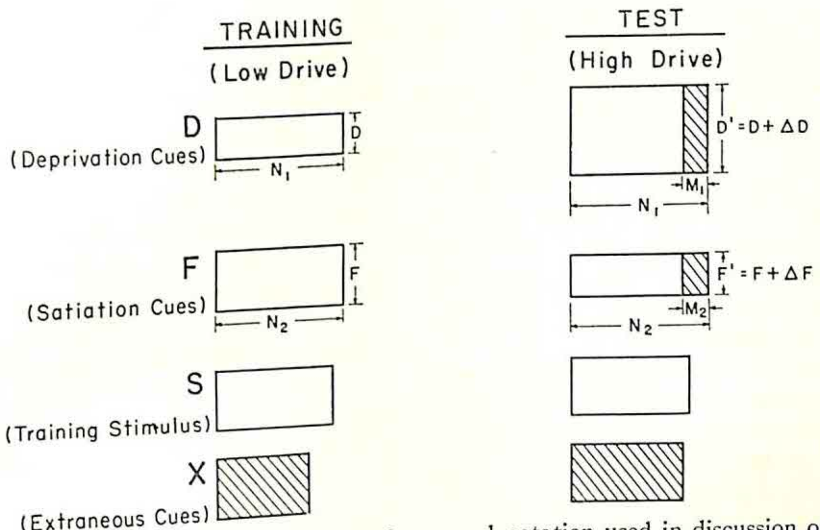

FIG. 1. Classes of cues in Estes' drive theory and notation used in discussion of effects of drive changes. (The width of a rectangle represents the number of cues, while the height represents the average sampling probability of those cues. For illustrative purposes, a drive increase is depicted.)

as stomach contractions or dryness of the throat; cues controlled or manipulated by the experimenter (the "stimulus," S), such as pitch or loudness of a tone; and extraneous cues (X), such as noises from the hallway. The weight of a class of cues is the product of the number of such cues available for sampling and the average sampling probability of those cues. The theoretical probability of a given response is equal to the proportion of all classes of cues conditioned to the response (weighted by the appropriate sampling probabilities) to the total weight of all cues available for sampling.

Figure 1 illustrates the changes associated with a drive increase between training and testing and presents the notation to be used in further discussion. D and F are the sampling probabilities of deprivation and satiation cues, respectively, before drive is changed, N_1 and N_2 are the numbers of deprivation and satiation cues available for sampling. D' and F' are the sampling probabilities of deprivation and satiation cues after the drive increase. Changes in the sampling probabilities are represented by ΔD and ΔF. M_1 and M_2 are the numbers of deprivation and satiation cues available for sampling which were not present during training. S and X are the total weights of the experimental and the extraneous cues. In Figure 1, the white areas represent the weights of cues conditioned to the response, while cross-hatched areas represent the weights of cues not conditioned to the response.

Assumptions Concerning the Effects of Drive. A few assumptions regarding the effects of drive changes on stimuli available for sampling and their sampling probabilities make possible rigorous deductions concerning the theoretical effects of drive changes on GSGs.

1. It is assumed that the average sampling probability of deprivation cues (D) increases monotonically and that of satiation cues (F) decreases monotonically with increased drive.

2. The deprivation and satiation cues available for sampling are as-

sumed to vary systematically with drive changes. If testing drive is different from training drive, some of both classes of cues which were available during training will be replaced by cues which were not available for sampling. To simplify deductions, it is assumed that the number of deprivation cues and the number of satiation cues available for sampling remain constant.

3. The combined weight of deprivation and satiation cues increases with increased drive and decreases with decreased drive, that is, $N_1\Delta D + N_2\Delta F$ is positive for increased drive, negative for decreased drive.

The following formulas follow directly from the above assumptions, which seem to underlie Estes' (1958) presentation.

Let G denote overlap with the training stimulus, with $G = 1$ denoting the training stimulus, $G = 0$ denoting a stimulus so remote that it has no overlap with the training stimulus, and $0 < G < 1$ denoting intermediate degrees of overlap. GS is the overlap of generalization stimuli with the training S. If, as in Figure 1, training has been carried to asymptote so that all stimulus elements except extraneous cues are conditioned to the responses, then the response probability at the end of training for any stimulus on the S continuum is:

$$P_{\text{train}} = \frac{N_1D + N_2F + GS}{N_1D + N_2F + S + X}. \qquad [1]$$

When testing is carried out under a different drive from that present during training, response probability is given by:

$$P_{\text{test}} = \frac{(N_1 - M_1)D' + (N_2 - M_2)F' + GS}{N_1D' + N_2F' + S + X}. \qquad [2]$$

In order to predict the effect of a drive change between training and testing on a GSG, we require a formula for $P_{\text{test}} - P_{\text{train}}$.

P_{test} can be written:

$$\frac{[N_1D + N_2F + GS] + [N_1\Delta D + N_2\Delta F - M_1D' - M_2F']}{[N_1D + N_2F + S + X] + [N_1\Delta D + N_2\Delta F]}.$$

Then

$$P_{\text{test}} - P_{\text{train}}$$

$$= \frac{\begin{array}{r}[N_1\Delta D + N_2\Delta F - M_1D' - M_2F'][N_1D + N_2F + S + X] \\ - [N_1\Delta D + N_2\Delta F][N_1D + N_2F + GS]\end{array}}{[N_1D + N_2F + S + X] + [N_1\Delta D + N_2\Delta F][N_1D + N_2F + S + X]}.$$

By expanding and grouping terms and letting $T = N_1D + N_2F + S + X$, we arrive at:

$$P_{\text{test}} - P_{\text{train}} = \frac{[S + X - GS][N_1\Delta D + N_2\Delta F] - [M_1D' + M_2F']T}{[T + N_1\Delta D + N_2\Delta F]T}. \qquad [3]$$

From Equation 3 a number of predictions can be made concerning the effects of drive changes on GSGs. Drive decreases will be considered separately from drive increases. Where predictions depend upon assumptions within the theory, they will be stated in the context of the relevant boundary conditions.

Lower Testing Drive than Training

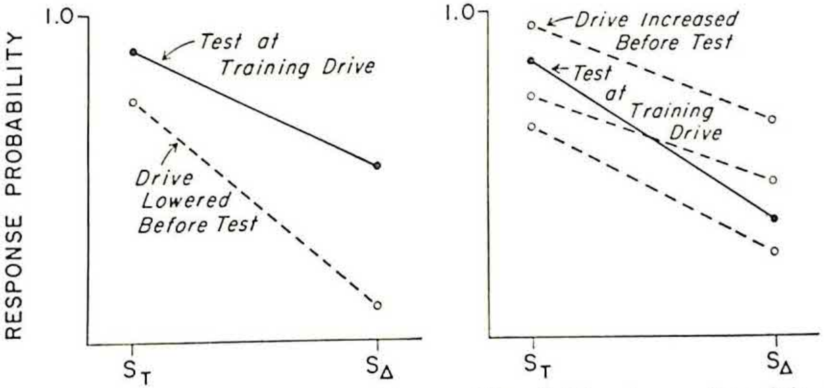

FIG. 2. Predicted effects of decreased drive on GSGs (left) and several possible effects of increased drive based on Estes' theory (right).

Drive. In this case $N_1\Delta D + N_2\Delta F$ is negative by assumption. The following predictions are derived from the formula for $P_{\text{test}} - P_{\text{train}}$.

1. The denominator in Equation 3 is always positive, since T is the total weight of all stimuli in the training situation, and the addition of $N_1\Delta D + N_2\Delta F$ to T yields the total weight of all stimuli in the test situation. Since $S + X - GS$ is also positive while $N_1\Delta D + N_2\Delta F$ is negative, and the positive expression, $(M_1D' + M_2F')T$, is preceded by a minus sign, the value of $P_{\text{test}} - P_{\text{train}}$ is always negative. Thus, the gradient is lower at S and at all GS.

2. The less the overlap of test stimuli with S, the lower the value of G and the higher the value of $S + X - GS$. Thus, at more remote stimuli, $P_{\text{test}} - P_{\text{train}}$ is a larger negative value than at the training stimulus. Hence, reduced drive steepens the GSG.

The lower and steeper gradients resulting from reducing drive between training and testing are illustrated in Figure 2.

Greater Testing Drive than Training Drive. By assumption, the value of $N_1\Delta D + N_2\Delta F$ is positive in this case. Several predictions follow:

1. At a given test stimulus, $P_{\text{test}} - P_{\text{train}}$ depends upon the relative weight of $N_1\Delta D + N_2\Delta F$ as compared to $M_1D' + M_2F'$; that is, essentially the change in sampling probabilities relative to the change in cues available for sampling. These two factors work in opposite directions with respect to the elevation of the GSG, $N_1\Delta D + N_2\Delta F$ tending to raise the GSG, $M_1D' + M_2F'$ tending to lower it. Thus, the GSG can be either raised or lowered by increased drive depending upon the values specified for these variables.

2. In Equation 3, for a given drive change, all quantities are constant over variations in the experimental stimulus with the exception of GS. Since $N_1\Delta D + N_2\Delta F$ is a positive value for increased drive, and $S + X - GS$ becomes a larger positive figure as G decreases, $P_{\text{test}} - P_{\text{train}}$ is greater at more remote stimuli than at the training S. Thus, drive increases lead to flatter GSGs.

3. If $P_{\text{test}} - P_{\text{train}}$ is negative at the training S, at remote stimuli $P_{\text{test}} - P_{\text{train}}$ may be positive since $S + X - GS$ is greater at remote stimuli. Thus, increased drive may decrease response probability at some stimuli and simultaneously increase response probability at other stimuli with less overlap with the training S. This "crossover" is more likely to occur if the major effect of increased drive

consists of changes in sampling probabilities (an effect represented in Equation 3 by $N_1\Delta D + N_2\Delta F$), and it is unlikely to occur if the major effect of increased drive is through changes in the drive cues available for sampling (represented by $M_1D' + M_2F'$). Again the prediction depends upon assumptions regarding the influence of $N_1\Delta D + N_2\Delta F$ relative to $M_1D' + M_2F'$.

Figure 2 illustrates several of the possible effects of increased drive including higher and flatter GSGs, flatter GSGs which cross the GSG at the training drive, and lower and flatter GSGs which do not cross the GSG at the training drive. Not all possible predictions are represented; for example, the noncrossover GSG below the GSG at the training drive could be higher at the training stimulus than the GSG which crossed.

A More General Formula. It will be recalled that in Figure 1 and in the formulas upon which the above predictions were based it was assumed that training had been carried to asymptote so that all but extraneous cues had been conditioned to the response. More general predictions which include the many situations in which training has not been carried to completion would be of interest.

To describe these more general conditions, some additional notation is needed. Let N_3 and S be the number and sampling probability of experimental cues and L_1, L_2, and L_3 be the numbers of deprivation, satiation, and experimental cues, respectively, which were not conditioned to the response at the end of training. M_1 and M_2 represent the numbers of drive and satiation cues which were conditioned to the response at the end of training, but were replaced by unconditioned cues as a result of a change in drive before testing. The formula for the change in gradients resulting from a change in drive is:

$$P_{\text{test}} - P_{\text{train}} = \frac{\begin{array}{r}(N_3S + X + L_1D + L_2F - (N_3 - L_3)GS][N_1\Delta D + N_2\Delta F] \\ - T[L_1\Delta D + L_2\Delta F] - T[M_1D' + M_2F']\end{array}}{[T + N_1\Delta D + N_2\Delta F]T}. \quad [4]$$

To some extent, the predictions from this formula differ from those stated above. However, it is again predicted that decreased drive will produce steeper GSGs, and increased drive, flatter GSGs. The steepening or flattening will be less marked the larger L_3 is relative to N_3.

Using this more general formula, the predictions concerning whether increased drive will raise or lower GSGs remain essentially the same as in Figure 2. The situation with respect to decreased drive is more complex and depends on assumptions about the different rates of change in the variables involved. Decreased drive may raise the GSG under some conditions rather than lower it. These conditions depend upon the values of N_1, N_2, N_3, L_1, L_2, L_3, M_1, M_2, etc., as specified in Formula 4.[2]

---

[2] For example, if the relative weight of deprivation and satiation cues conditioned to the response is initially less than the relative weight of experimental and extraneous cues conditioned to the response, and if there is no change in drive cues available for sampling and no change in the weight of drive cues conditioned to the response relative to the weight of all drive cues, the probability of the response will increase with decreased drive. This is because the total weight of drive cues

If as a result of reduced drive the GSG is raised at the training stimulus, a crossover effect would be predicted for larger absolute values of $N_1D + N_2F$ relative to $M_1D' + M_2F'$.

THE RESPONSE ACTIVATING OR ENERGIZING ACCOUNTS OF DRIVE

In this section we will consider a family of theoretical accounts which interprets the effect of drive on stimulus generalization primarily in terms of a multiplicative effect of drive on the habit strength of a response (e.g., Broen et al., 1963; Hull, 1952, Ch. 8; Rosenbaum, 1951, 1953; Spence, 1958). Although these accounts are based on the Hullian treatment of drive as a response energizer, none conforms strictly to the Hullian postulate system, which includes statements regarding the effects of changes in drive-stimulus intensity on response strength (Hull, 1952, Corollary 12). Hull was dissatisfied with his treatment of stimulus intensity (1952, pp. 89, 220) and chose to illustrate the effect of changed drive on a gradient of stimulus generalization with computations based solely on the energizing aspect of drive (p. 225). The other writers cited above have followed suit. In the family of accounts discussed here the possible effects of the stimulus aspect of drive are not denied. However, since these accounts relate *SG* only to the energizing aspect of drive, they imply that the major effects of drive on *SG* can be accounted for in terms of the energizing function of drive.

The accounts of the relationship between drive and *SG* which emphasize the energizing aspect of drive are based on, or seem to imply, the following theoretical specifications:

1. When a response has been trained to a stimulus, the habit strength (H) of the response will generalize to other stimuli, with the generalized H being a monotonic decreasing function of stimulus dissimilarity.
2. The same stimulus may evoke more than one habitual tendency to respond (e.g., H_1, a dominant response, and H_2, a competing response).
3. At any stimulus, drive (D) multiplies the habit strength of all tendencies to respond elicited by that stimulus.
4. With other variables held constant, reaction potential (E) is the product of $D \times H$, and thus $E_1 - E_2 = D(H_1 - H_2)$. E oscillates in strength from moment to moment, the momentary Es being distributed approximately normally. Competing responses have equal standard deviations of momentary Es resulting from oscillation, and competing Es oscillate asynchronously. Despite varying statements regarding oscillation in the different accounts, the above statement reflects computing practices which have been quite consistent.
5. In order for a response to occur, at the time of testing its momentary E must be stronger than the momentary Es of competing responses, and must be stronger than a threshold value which is constant and above zero.

Energizing Accounts and GSGs

In discussing implications of the above account, the probability of a response (P_R) will be used as the dependent variable because of ease and directness of derivation from the

---

decreases as drive decreases (see Assumption 3). However, if the total weight of drive cues decreases very slightly and the number of deprivation and satiation cues available for sampling changes substantially, the probability of the response will decrease with decreased drive.

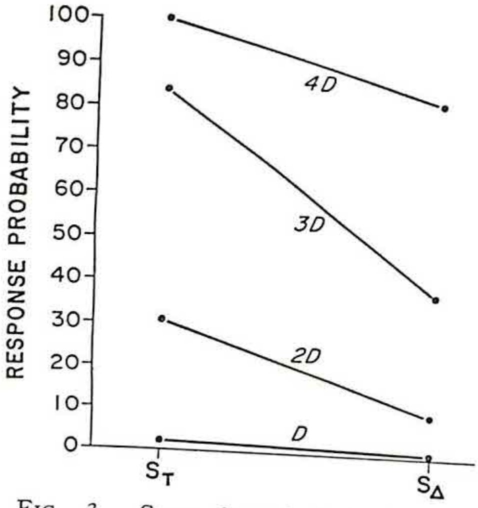

FIG. 3. Some hypothetical GSGs at different drive levels in a situation without competing responses computed according to Spence's theory. (Time-limited trials which are brief enough to rule out within-trial E variation are assumed. Longer trials would give higher response probabilities. Under the lowest drive level illustrated, D, $E_1 - L = .5$ at S_T and .1 at S_Δ. Threshold = 1.0. These values are in terms of the standard deviation of momentary E; E minus oscillation.)

energizing account and comparability with derivations from Estes' account.

Noncompetitional Situations. Let us first consider the simple situation where one R is trained and no competing Rs are present. Most discussions of the energizing aspect of drive and its relationship to SG reflect the implications of the energizing account for this situation which are illustrated in Figure 3. The experimental situation represented is one in which a response has been trained to the stimulus S_T following which observations are made on the first test trial at either S_T or separate presentation of a generalization stimulus, S_Δ. Thus, the response probabilities in Figure 3 are simply the proportions of the distribution of momentary Es which are above threshold at a specific level of drive.

Consistent with the implications illustrated in Figure 3, most writers using this account (Dollard & Miller, 1950, p. 53; Hull, 1952, p. 225; Knopf & Fager, 1959; Mednick, 1958; Rosenbaum, 1953; Spence, 1958) have stated that at higher levels of D, GSGs will be elevated (increased probability of R). Steeper GSGs under high D than low D (see Figure 3—D, $2D$, $3D$) are also expected (Rosenbaum, 1953, p. 35; Spence, 1958, pp. 83, 84), although it is also noted that this GSG steepening may not be observed when certain dependent variables are used. The expected flattening of GSGs under higher D when dependent variables are near their upper limits at S_T (compare $4D$ with $3D$ in Figure 3) has also been noted (Rosenbaum, 1953, p. 4).

Competitional Situations. When stimuli evoke competing responses and the E distribution of R_1 is entirely above threshold, the predictions are substantially the same as in the noncompeting situation represented in Figure 3. Higher D leads to higher and steeper GSGs, with flattening as P_{R_1} at S_T approaches 1.00; lower D leads to lower and flatter GSGs unless P_{R_1} at S_T was initially close to 1.00, in which case some steepening will precede the flattening. A difference from the noncompetitional situation is that at remote stimuli where a competing R is dominant, P_{R_1} will be reduced by drive increases and raised by drive decreases.

When a portion of the E distribution of R_1 is below threshold (L), the effects of competing responses interact with the threshold to yield predictions which are considerably more complicated. We shall present the results of computations which illustrate the diversity of the possible predictions, some of which are considerably different from predictions which have been made in the available accounts of D

and *SG* which are represented in Figure 3. Since the predictions suitable for a specific situation depend upon the boundary conditions, it is imperative that predictions and conclusions be presented in the context of the relevant boundary conditions, as we shall attempt to do.

The experimental situation we will use to illustrate these varied effects is the same as the situation represented in Figure 3 with two important differences. First, presentation of either S_T or S_Δ is assumed to evoke both R_1 and, at lesser H, a response (R_2) which is incompatible with R_1. The competing response may be the result of any of a number of factors, such as previous training to S_T or S_Δ or generalization from another stimulus. Second, trials are ended only by the occurrence of either R_1 or R_2. The major implications of the energizing account in this situation are illustrated in Figure 4. The computations were based on Spence's (1956, Appendix A) method for computing response probability in a situation with two alternate responses. GSGs have been computed for one level of drive (represented as D) and higher levels of drive, each twice that of the next lower level ($2D$, $4D$, $8D$).

As Figure 4 illustrates, when a competing R is present, either increased or decreased drive may elevate the GSG while also sharpening or flattening the GSG, or it may lower the GSG with concurrent sharpening or flattening. Two additional possibilities not shown in Figure 4 should also be noted. Increased D may raise the GSG at S_T while lowering it at a generalization stimulus (e.g., if $H_1 < H_2$ at S_Δ). Alternatively, it is possible to find conditions where a decrement in P_{R_1} at S_T and an increase in P_{R_1} at S_Δ would be expected under increased D—for example, when $L = 1$, $E_1 - L = 1.5$, and $E_1 - E_2 = 1.5$ at S_T, while at S_Δ, $E_1 - L = 1.2$, and $E_1 - E_2 = .2$. (When D is doubled, these conditions would lead to a very small R_1 decrement at S_T and facilitation at S_Δ.)

Dependence upon Boundary Conditions. Since such varied and contradictory results are all theoretically

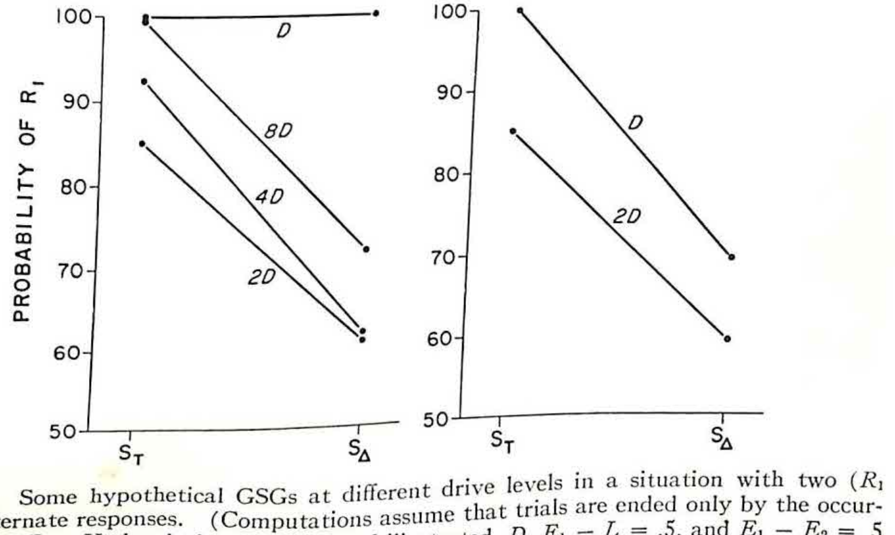

FIG. 4. Some hypothetical GSGs at different drive levels in a situation with two (R_1 and R_2) alternate responses. (Computations assume that trials are ended only by the occurrence of R_1 or R_2. Under the lowest drive level illustrated, D, $E_1 - L = .5$, and $E_1 - E_2 = .5$ at S_T. At left, $E_1 - L = .1$, and $E_1 - E_2 = .1$ at S_Δ under D. At right, $E_1 - L = .2$, and $E_1 - E_2 = .1$ at S_Δ under D. Threshold $= 1.0$.)

possible, it would seem that in order for the energizing account to be most useful or testable, predictions from the account should be made in conjunction with statements of boundary conditions which would make the predictions theoretically necessary. This is especially true when competing responses are evoked in a situation or where predictions are made about GSG slope. Predictions for GSG elevation in noncompetitional situations are monotonic (e.g., lowered GSGs are not possible with increased D, and GSG elevation cannot occur under decreased D) and are consistent with the predictions which have usually been made from the energizing account.

A number of variables must be considered in determining the boundary conditions which make specific predictions possible. The following are some implications of the energizing account for GSGs, presented in the context of the relevant boundary conditions:

1. One of the more unusual implications is that a GSG may be lowered by increased drive or raised by decreased drive. These inverse relationships between D and GSG elevations are possible only if trials can be ended solely by occurrence of an experimenter-designated dominant or competing response (as in Figure 4) rather than being time limited. Thus, if trials have a constant length and R_1 is dominant, according to the energizing accounts P_{R_1} cannot be reduced by increased drive (see Broen & Storms, 1961, for further discussion). Thus, the use of time-limited trials may be desirable in experimental tests to reduce the ambiguity of the relevant predictions.

2. The above-described threshold effects which may result in lower P_{R_1} under higher D are at a maximum when $E_1 - L$ and $E_1 - E_2$ are small (see Spence, 1956, Figure 71). In such a situation, the proportion of the distribution of E_2 raised above threshold by D is at its greatest as compared to the counteracting effect of drive in increasing $E_1 - L$ and $E_1 - E_2$. A testable consequence is that if increased D raises the GSG at any stimulus, further increases in D (acting in a situation with even greater $E_1 - L$ and $E_1 - E_2$) will have a monotonically increasing relationship to the height of the GSG at that stimulus. This holds for both competitional and noncompetitional (see Figure 3) situations. The converse, that if lowered D lowers the GSG, further decreases in D will have a monotonically decreasing relationship to the height of the GSG, is necessarily true only for noncompetitional situations.

3. Specific predictions about whether flattening or sharpening in gradients of response probability will occur under increased or decreased D are very difficult to make from the energizing account. For example, in th situation illustrated in Figure knowledge of $E_1 - L$ and $E_1 - E_2$ both S_T and S_Δ and amount increase in D is needed to make suc predictions. Even if ingenious pri experimentation were done to obt knowledge of the level of L for experimental situation, additional assumptions would be needed to translate most experimental drive variation (e.g., a change in shock intensity) int a specific change in the theoretical construct D. This kind of specific knowledge is needed to make predictions because of the nonmonotonic relationships between D and slope of GSG which are implied by the energizing account.

Drive Stimuli and GSGs

Drive-stimulus (S_D) effects have been omitted in the above discussion of energizing accounts of drive and GSGs. However, both Hull (1952) and Spence (1956) have mentioned the importance of S_D contributions to H and E. Therefore, we will comment briefly regarding effects of S_D on GSGs and note some implications and problems. Rather than include competing responses specifically associated with different intensities of S_D, we shall limit our discussion to the effects of S_D generalization.

When training is carried out at one drive level and testing at another, the stimulus situation at the time of testing is less similar than it would be if training and testing drives were the same. Thus, the H associated with S_D would be reduced, with a corresponding reduction in E.

Possibly because of difficulties in the theoretical formulation and empirical measurement of S_D or his greater interest in energizing effects, Spence has generally controlled for S_D effects or omitted them from sideration (Spence, 1958; Spence, drich, & Ross, 1959). Although ncerned with difficulties in his rmulations, as we shall see, Hull 952) provides some specification of effects in his postulate system:

> h drive condition (C_D) generates a characteristic drive stimulus (S_D) which is a monotonic increasing function of this state [p. 7, Postulate V].

> When a habit is set up in association with a iven drive intensity (S_D) and its strength is tested under a different drive intensity, there will result a falling gradient of ${}_S\bar{H}_R$ and ${}_S\bar{E}_R$ [p. 11, Corollary xii].

This statement that ${}_S\bar{E}_R$ will fall off implies that any change in drive will result in a lower GSG as compared with training drive. Perhaps Hull intended that only the contribution of S_D to E would be lowered, so that the corollary would apply only to situations where, as the result of learning, drive stimuli have come to have primary control over the response. If other than drive stimuli contribute substantially to H, the energizing effect of D could outweigh the drive-stimulus decrement, leading to increased E and higher GSGs, as Hull (1952) predicts in Theorem 59 (p. 226).

There are other problems in the application of Hull's principles concerning S_D:

1. Since S_D vary in intensity, S_D generalization follows the principles of stimulus-intensity generalization (Hull, 1952, p. 11). However, Hull (1952, p. 89) cites evidence which he says implies a defect in his analysis of stimulus-intensity generalization, and therefore has substituted the methods used for qualitative-generalization gradients rather than apply his methods for computing stimulus-intensity generalization gradients (p. 220).

2. Since Hull states only that S_D is a monotonically increasing function of C_D, it is impossible to compute the amount of S_D change associated with a given change in D.

3. The energizing effect of stimulus intensity, V, is included in the formulas for stimulus-intensity generalization. In derivations regarding S_D generalization, are V and D to be considered separately in the computations, or should V be omitted since the same conditions, C_D, give rise to both? If one shares Hull's lack of confidence in his stimulus-intensity generalization analysis, this question may be unimportant, except insofar as answers contribute to more adequate formulations.

As we have seen, almost any effect is consistent with the energizing account, given the proper conditions. The introduction of substantial S_D effects leads to modifications in the predictions and even greater complexity in the boundary conditions which have to be considered. Since a clear theoretical statement regarding S_D effects is lacking and the evidence concerning S_D changes and GSGs is sparse, we shall avoid compounding the complexity of the predictions and go on to consider one clear implication of Hull's account of the effects of S_D changes.

Since the same drive level is present at all stimuli on the stimulus continuum, when testing drive is different from training drive, the same E is withdrawn at each point on the stimulus continuum. Employing the formula for the withdrawal of E (Hull, 1952, p. 9), we find that removing a given amount of E will reduce the total E by a larger amount when the initial E is small (as at generalization stimuli) than when the initial E is larger (as at the training stimulus). Let E_1 represent the E at the training stimulus, E_2 at a generalization stimulus, and E_3 the amount to be withdrawn because of a drive-stimulus change, while M is the maximum possible E. The difference in the changes in E_1 and E_2, $(E_1-E_1')-(E_2-E_2')$, is

$$(E_1 - E_2)\left(1 - \frac{M}{M - E_3}\right).$$

Since this value is always negative, the change in the smaller E (E_2) is greater than the change in the larger E (E_1). Hence, in terms of E, the resulting GSG is steeper.

Thus, ignoring the energizing effects of drive, S_D generalization invariably leads to lower and steeper GSGs in terms of E. (In terms of response probabilities, some flattening would be introduced if the E_1 distribution moves below threshold.) This is the one invariant prediction regarding D and SG which follows from Hull's theory. It holds whether S_D generalization follows stimulus-intensity or qualitative-generalization principles. It is also independent of which positions are taken regarding the other problems in Hull's S_D formulation. However, these problems affect the manner in which S_D effects are combined with energizing effects to modify the predictions of the energizing accounts.

COMPARISON OF THE THEORIES

An important feature of $D \times H$ energizing accounts of drive and SG is that they sometimes predict nonmonotonic relationships, for example, sharpening of GSGs followed by flattening with increased D. Thus, specification of amount of D is necessary to arrive at a prediction for a given experimental situation. Although this feature makes experimental tests difficult, drive scaling procedures such as those of Pinneo (1961) may help make specific predictions possible.

Additional difficulties are involved in making predictions which depend upon both an energizing effect and a drive-stimulus change effect of drive or upon changes in sampling probabilities and drive-stimulus elements. Shared by the theories under consideration, these difficulties have to do with the fact that differences in drive stimuli must be inferred from the very drive-inducing conditions and consequent changes in dependent variables which are used as evidence of a change in drive. Predictions often depend upon the relative predominance of energizing and drive-stimulus effects or sampling probability and drive-stimulus change effects, while the determination of this predominance in an experimental situation is at least extremely difficult. Determination of drive-stimulus intensities as a function of D and the psychophysical scaling of drive intensities would be required for the application of Hull's formulas, and similar specification would be required for other $D \times H$ accounts. Similarly, the number of drive-stimulus elements and the change in the elements available for sampling for a given drive change are needed to make specific predictions, but are not specified in Estes' theory and cannot be derived from empirical determinations.

Although most predictions are dependent upon specification of amount of initial drive, change in drive, and change in drive-stimulus intensity or drive-stimulus elements, some predictions from Estes are more general than any from the $D \times H$ accounts. Estes' theory leads to at least two unequivocal predictions. The first is that increased drive will lead to flatter GSGs while decreased drive will bring about steeper GSGs. Since this prediction is independent of experimental arrangements, Estes' theory is more testable than the others in this respect, and testability is certainly a desirable feature of a theory. Another unequivocal prediction from Estes is that a response will have some nonzero probability of occurrence even at extremely remote generalization stimuli (because some drive stimuli are conditioned to the response). If any stimulus situation whatever can be found such that $p = 0$, this prediction is disconfirmed. Competing responses and the threshold allow Hull and Spence to predict $p = 0$ in some circumstances. A recent revision of Estes' theory (1959a) stating that responses are conditioned to patterns of stimuli rather than to individual stimulus elements circumvents this difficulty, but at the expense of greatly increased complexity and decreased testability.

As we have seen, the combination of threshold effects with response competition in $D \times H$ accounts leads to very complex and nonmonotonic relationships between D and SG. These relationships are not easily compared with derivations from Estes. However, clearer predictions not requiring quantitative specification of variables can be made for classes of situations in which E_1 is high enough to be above threshold or where there is no response competition. These predictions can be compared with the corresponding predictions from Estes.

The predictions from Hull and Spence differ from those derived from Estes most strikingly with regard to slope, where Estes' predictions are clear-cut, as we have seen, and those from Hull and Spence are ambiguous. However, if energizing effects predominate, with R_1 suprathreshold and P_{R_1} not initially close to 100%, the prediction from Hull or Spence is that decreased drive will lead to flatter GSGs, which is more consistent with the meager evidence (Dinsmoor, 1952; Newman, 1955; Thomas & King, 1959) than Estes' unambiguous prediction of steeper GSGs with lowered drive. With higher testing than training drive, some evidence shows steepening (Dinsmoor, 1952; Newman, 1955) and some shows flattening (Broen et al., 1963). In the Jenkins et al. (1958) experiment, which controlled for drive-stimulus changes by training and testing at the same drive, higher drive was associated with flatter GSGs, as predicted by Estes. However, in a portion of the experiment by Thomas and King (1959) which controlled for drive-stimulus changes in the same manner, higher drive was associated with steeper GSGs, and for drive levels higher than

Psychological Review
1966, Vol. 73, No. 2, 128–150

MECHANISM OF THE FIGURAL AFTEREFFECTS

LEO GANZ

University of California, Riverside

It is suggested that the figural aftereffects are actually a species of simultaneous illusion in which the afterimage of the inspection object acts as an inducing figure. The afterimage displaces the test contour in phenomenological space because the inducing contour exerts inhibition on the test contour. A simple mathematical theory is described which accounts for the spatial distribution of displacements.

When contours of visual objects are close together, they often induce shifts in their respective *apparent* locations. Since the objects are not, in fact, changed in their positions in the visual field, it must be the neural correlates of these edges or lines which somehow interact, shifting each other's positions in their topographic projection in the visual system. The phenomenological result is a deviation in perception from veridicality, an illusion. In Figure 1 the lines of the square on the left traversing the concentric circles are so affected by each intersecting arc as to increase slightly the angle of each intersection, especially if the angle of intersection is somewhere midway between zero and 90°. This gives the square a slightly concave appearance. Many illusions—such as those of Hering, Wundt, Müller-Lyer, Zöllner, and Poggendorff—seem to involve such a common principle.

There is another class of contour shifts—the figural aftereffects—in which the inducing contour[1] seems to act proactively. The experimenter (*E*) presents an inducing figure in the observer's (*O*'s) visual field, then removes it, and then presents a test figure. Even with this temporal separation between the two figures, the first still continues to act on the second so as to shift the second's apparent position.

Figure 2 illustrates an experiment involving such a procedure (Köhler & Wallach, 1944). *O* is first shown the two black rectangles (marked I in the figure), which are the inducing figures. He fixates the X mark. Then, usually after about 60 seconds of inspection, these rectangles are removed. Now the two outline squares (each a test figure) are shown, and *O* again fixates the X mark. Frequently *O* will report that the square on the left looks higher than that on the right (in fact they are equally high). This is a typical *displacement effect;* contours are repulsed away from the edge of the previously shown inspection figure. The displacement effect is typically maximal not when the inducing and test lines abut, but when they are actually

---

[1] Using commonly accepted terminology, we call the *inducing figure* a line, edge, object, etc., introduced as an independent variable into *O*'s sensory field, visual or other. The test figure is a dependent variable, a type of probe stimulus, upon which the effect is exerted. In this paper we will be concerned mostly with a dimming or displacement in appearance. The *comparison figure* is any figure which is placed far enough away from the inducing figure so as to be only mildly affected by the independent variable. Thus, any change in appearance or position of the test figure is gauged by contrasting it with the less affected comparison figure. The *interfigural distance* (*S*) is always the distance between the inducing and the test figure (in minutes of visual angle unless otherwise noted).

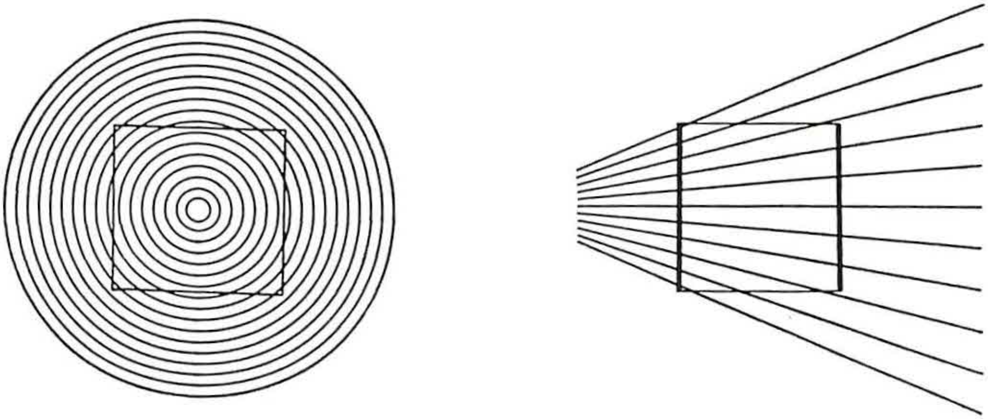

FIG. 1. Two geometrical illusions. (In the left figure, the lines of the square traversing the concentric circles are seen by many observers to turn toward the center of the circles. In the right figure, the square is seen as a trapezoid, the side closer to the intersection of the converging lines looking larger. Note that in both cases the phenomenon involves a growth in the angles formed by lines intersecting at 10°–40°.)

separated. Köhler and Wallach have called this the paradoxical distance effect.

In the experiment illustrated in Figure 3, we first present *O* with a circle as an inducing figure, and *O* fixates the X mark on the right edge of the circle. This figure is then removed, and an outline rectangle, T, is presented, with the same fixation mark, X, as shown. *O* reports that the left part of the rectangle looks as though it has shrunk (a *size effect,* often equivalent to the displacement effect and again involving a repulsion away from the inspection-figure contour). Subjects also report that the test contours left of the fixation mark (inside the former inspection circle) look dimmer, or washed out, with less contrast (the so-called *color effect*). Lastly, the left half of the rectangle is often seen further back in space (the *depth effect*).

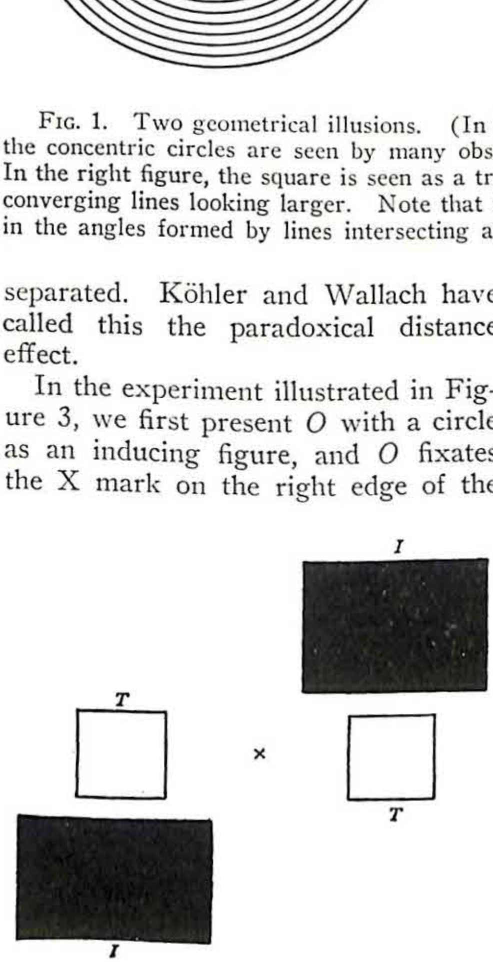

FIG. 2. A stimulus configuration used by Köhler and Wallach (1944) to study figural aftereffects. (*O* fixates the X mark, and the I figure is presented. Then the I figure is removed, and the T figures are presented. The T figures are then seen more dimly, and repelled away from the former locus of the I figures. The repulsion is reported by many *O*s when I and T are presented together.)

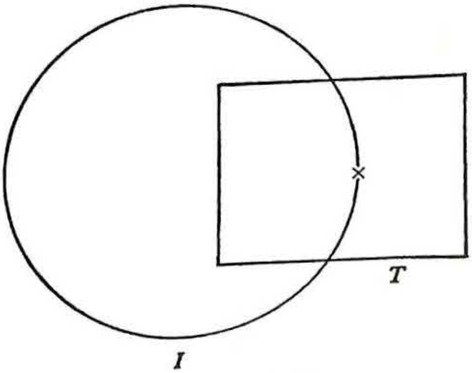

FIG. 3. Same as in Figure 2.

It is not necessary to present the inducing and test figures in succession to see the displacement effects. Many *O*s looking at Figures 2 or 3 from about 3 meters distance, with the inducing and test figures shown simultaneously, still report the displacement effects we have described. This similarity in results between the simultaneous and successive presentations was already apparent to Köhler and Wallach in their earlier experiments and has recently been reiterated (Logan, 1960). Do the two sets of displacement effects have identical mechanisms? This constitutes the point of departure of the present theory.

The initial theory of the "Figural After-Effects" (Köhler & Wallach, 1944) used simple physical principles involving current flow in a volume conductor such as the visual cortex to derive the distortions in perceptual appearance. The theory attracted considerable attention, some of it critical (Day, Pollack, & Seagrim, 1959; Hebb, 1949; McEwen, 1958; Spitz, 1958). The main criticisms have been that:

1. Lateral direct-current flow can be disrupted without noticeable changes in perceptual function (Lashley, Chow, & Semmes, 1951; Sperry & Miner, 1955; Sperry, Miner, & Myers, 1955).

2. The theory assumes that a topologically faithful "picture" is transmitted to the cortex (Hebb, 1949), but in fact each synaptic stage is the site of radical information processing (Hu-

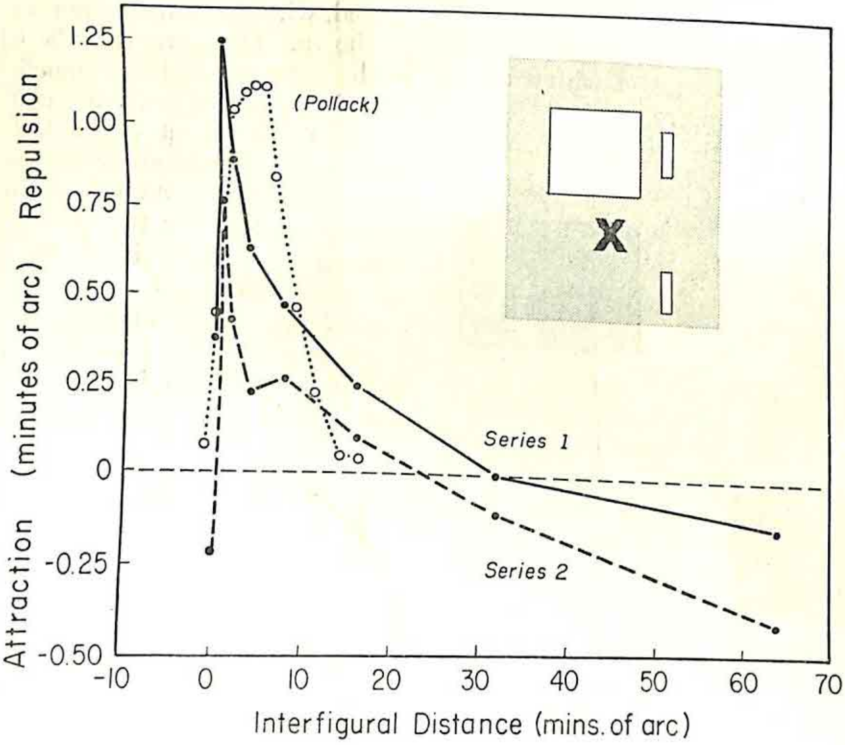

FIG. 4. The magnitude of the displacement effect when the distance between the inducing and test figure is varied. (The inducing figure is a white square and the test figure, a white vertical line. *They are presented together,* as depicted in the insert, in the manner of a simultaneous illusion. The curves on the left represent two successive series of measurements with I and T figures presented together, $N = 60$. The rightward curve, Pollack, 1958, represents data, adjusted in height, from *a successive presentation:* I and then T.)

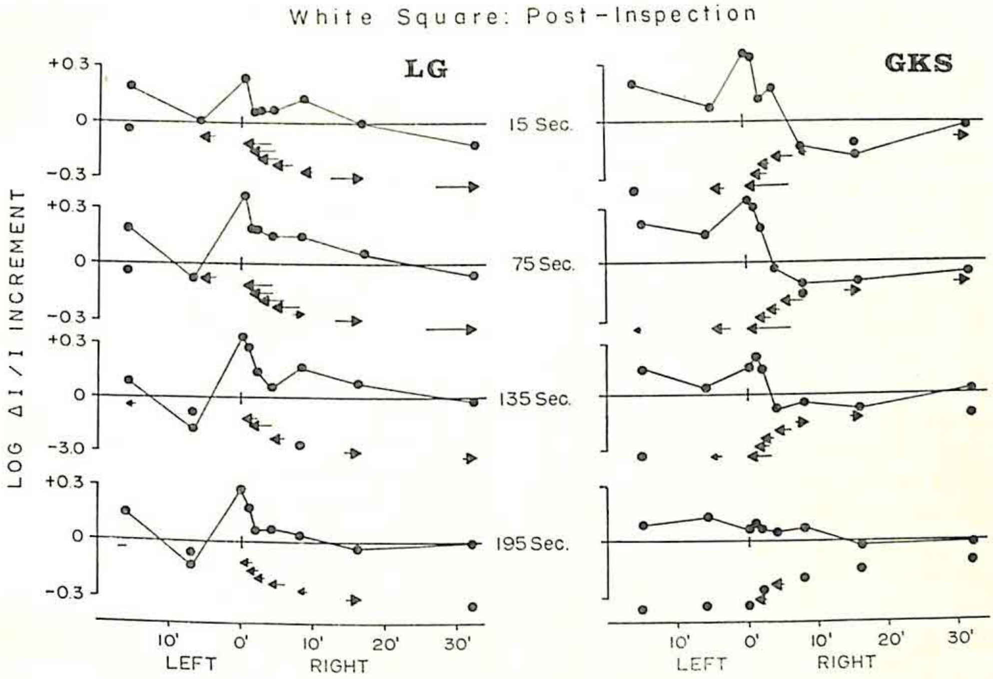

FIG. 5. The solid lines represent the *changes* in the detection of a vertical white line at various distances from the edge of a previously presented white square *after* a 2-minute inspection of the square. (Positive values represent a loss in sensitivity. The arrowheads point in the direction of contour displacements at those same distances. Measurements are depicted for 1-minute intervals—15, 75 seconds, etc. Left of the origin, the vertical white line was *inside* the locus of the previously presented square, $N = 2$. Arrows pointing toward the origin, 0′, represent attraction effects.)

bel & Wiesel, 1959, 1962; Kuffler, 1953; Lettvin, Maturana, McCulloch, & Pitts, 1959).

3. The theory is basically a fatigue theory, yet the effects build up almost instantaneously (Ikeda & Obonai, 1953; Köhler & Wallach, 1944, p. 355; Oyama, 1953, 1956). In fact, as we have seen, one often obtains quite comparable results by presenting the inducing and test figures simultaneously. In a recent experiment, 60 subjects were shown a simultaneous presentation of the inducing and test figures originally used by Köhler and Wallach. The contour repulsions obtained are given in Figure 4. A comparison with an experiment in which the figures were presented successively (Pollack, 1958) suggests a close correspondence of the two sets of results (see also Ikeda & Obonai, 1955; Ikuta, 1956; Logan, 1960; Nozawa, 1953).

4. Displacements sometimes occur *toward* satiated areas, as in the experimental results shown in Figure 5 (see also Ganz, 1964; Ganz & Day, in press; Nozawa, 1953a, 1953b; Smith, 1954).

5. Independent indexes of the hypothetical construct "satiation"—for example the color effects and the displacement effects—can be shown to be independent (Ganz & Day, in press). It is likely that they can occur at different neural loci.

The Osgood-Heyer (1952) theory is also a fatigue theory and so cannot be reconciled with Criticisms 2–5 above. As a hypothetical construct, it has not accorded well with some facts of visual physiology (Deutsch, 1956).

A NEW THEORY OF FIGURAL AFTEREFFECTS

To provide a satisfactory mechanism of the figural aftereffects, two phenomena must be accounted for *independently:* (*a*) How do contours displace one another? and (*b*) How do contours interact when they are temporally separated? We will first provide a somewhat general and qualitative description of a theory. This will be followed by a more detailed exposition.

It appears that most of the difficulties inherent in the earlier theories evolved from their attempt to deal with both the displacement effect and the time-binding (proaction) effect with a single mechanism. Since the two effects display a rich variety of independent behavior, there is no recourse but to consider two separate mechanisms: The two chosen in the present study are lateral inhibition (*I*) and light adaptation. We want to explore the possibility that all *contour displacements* which are induced by neighboring figures and are restricted to about 10 or 20 minutes of arc in human perception are due to the operation of such inhibition. This is considered to apply to geometrical illusions (such as in Figure 1) and to many figural aftereffects (such as Figure 2 or 3). In this, our theory owes something to the ideas of Day (1962) and Deutsch (1964) in which similar approaches were suggested. As we will show below, the inhibition typically diminishes as the distance between the contours is increased. This descending function is actually steepest at very small distances, just a few minutes of arc. It is easy to understand that when a contour is being inhibited by such a steep function, the portion of the contour near the inhibiting figure will be inhibited more than the portion at a greater distance. This must shift the mean of the contour's distribution *away* from the inhibiting figure. This is the repulsion effect.

Now such a theory, by itself, would lead one to think that the greatest repulsions occur when the two contours are *least* separated, which is when they touch. Yet the distance paradox tells us that empirically it has been found that repulsions are small or nil when the two contours touch. There are at least two plausible reasons why this should be so. One is that at very small interfigural distances, the two contours summate their excitations. In such a situation, it is likely that *O* actually confuses the inducing-figure and the test-figure contours (Taylor, 1962). Secondly, it is known that eye movements are present during the presentation of the test figure. Such movements must sometimes cause the test figure to be placed on one side of the inducing figure's edge and sometimes on the other. This is increasingly the case as the interfigural separation is made smaller. The displacement effect, a repulsion, is opposite in direction on the two sides of the inducing figure. Thus, when such crossovers are frequent, we will be adding judgments made now with one displacement, now with an opposite one, with the net result close to zero. Thus, net displacement (which is what the grouped experimental results give us) will approach zero as interfigural distance approaches zero. Although we are in general agreement with Taylor that contour confusion must play a part, the evidence below suggests that this is not sufficient. The present model explores the second alternative.

We next consider the temporal problem: How does the inducing figure act proactively, crossing a time gap to displace the test figure? When *O* views a dark figure, there is a corresponding retinal area which dark adapts more than its surround; the portion of the retina of a subject where a bright figure is imaged will be a site of greater light adaptation in comparison with the surround. When the inducing figure is then removed, *O* will view stimuli with a heterogeneously adapted retina: a retina with some regions that are more dark adapted and regions that are more light adapted than the remainder. The result is that under normal levels of illumination, the subject will, in fact, continue to perceive the inducing figure. It will be seen as a light-dark reversal of the original, and it will decay in time. But while it is present, its edges will interact with the test figure as they do in simultaneous illusions. This is because the stimuli which are created by a heterogeneously light-adapted retina are stimuli which in many respects *are identical*, from a neural point of view, to those created by a distal stimulus onto a homogeneously adapted retina. In many situations the subject cannot actually discriminate between the two inducing figures, an ordinary stimulus originating in the distal environment and an afterimage (Barlow & Sparrock, 1964; Brindley, 1963). We suppose that the contours interact, even when interocularly produced, at the visual cortex to produce simultaneous illusions. The afterimage exists beyond the preservation of the distal figure and decays in time. This is the mechanism which bridges the time gap in the figural aftereffect situation; or, more properly, the processes of light and dark adaptation bridge that time gap.

Below we first examine the evidence on lateral inhibition, and we then present the theory in a quantitative form. We next examine the spatial distribution of displacement effects as they follow from the theory. In a subsequent paper we will compare the growth and decay of light and dark adaptation on the one hand and the growth and decay of figural aftereffects on the other. The similarities in temporal properties, which are quite striking, will be used to substantiate our argument.

Lateral Inhibition

Psychophysical Evidence. Because of diffraction at the edge of the pupil and because light is scattered within the eye (Bartley, 1951, pp. 924–926; Boynton & Riggs, 1951), slit-shaped stimuli of high contrast are projected on the retina as broad, diffuse distributions (Flamant, 1955) considerably degraded in contrast. The anatomy and physiology of the retina suggest considerable spatial integration, even in the fovea, extending minimally to many minutes of arc and maximally (as in the periphery) to 1 or 2 degrees of arc, perhaps more. These facts might seem difficult to reconcile with the finding that in many vertebrate species visual resolution (the smallest consistently detected gap between two thin bars) is of the order of a single minute of arc. In fact, it is well known that the effects of spatial integration are held in check by a counterforce: lateral inhibition. Psychophysically, lateral inhibition is manifested by the appearance of contrast: Black solid figures appear blackest at their edges with white; white surfaces appear brightest at their edges with dark figures. Phenomenological contrast was known to Aristotle, was utilized technically by Leonardo da Vinci to enhance light values (Boring, 1942), and was first examined quantitatively by Lehmann (1886) and Hess and Pretori (1894).

Of the aspects of simultaneous contrast which are of interest to us, it should be noted that brighter stimuli inhibit dim ones more readily the greater the brightness difference of the two patches. This can be seen in Fig-

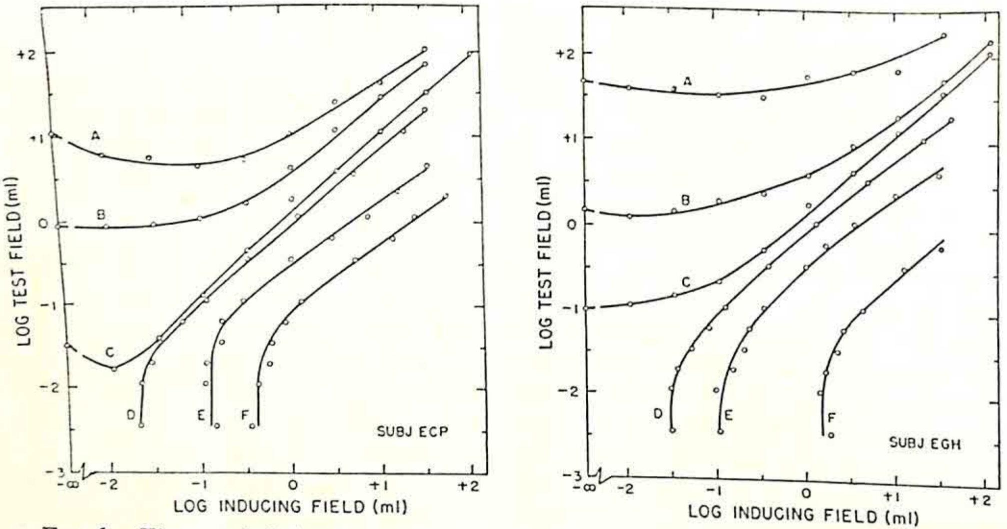

FIG. 6. The test-field luminance needed to match a constant comparison field. (When the inducing field, which surrounds the test field, is made more intense, it inhibits the action of the test field. The test field must then be increased in luminance to obtain a match with the comparison field. A, B, C, etc., indicate decreasing intensities of the comparison field—Heinemann, 1955.)

ure 6, taken from Heinemann (1955). In the figure the more intense inducing stimulus produces more inhibition. And for a range of values, it appears that the log matching luminance is a linear function of the log inducing luminance. One study (Leibowitz, Mote, & Thurlow, 1953) reports the dimming of the test field by the inducing field to be proportional to the log inducing field luminance.

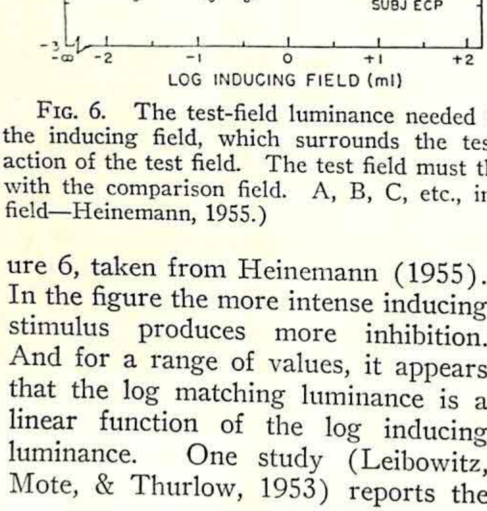

FIG. 7. The amount of inhibition exerted by a figure on another when their separation is varied. (The various functions depict effects of inducing figures of different intensities. An ordinate value of 1.00 log mL represents a 10-fold dimming of the test figure; zero mL represents no effect. Adapted from Leibowitz et al., 1953.)

Lateral inhibition is quite sensitive to the distance between two stimuli. For example, Beitel (1936) found that the empirical curve relating the amount of inhibition of the inducing stimulus on the test stimulus, as a function of luminance, was steeper when the two stimuli were adjoined more closely. In Figure 7, adapted from the data of Leibowitz et al. (1953), we see that inhibition diminishes when the distance between two stimuli is increased. *Most of the inhibition occurs within the initial 10 minutes of arc, whether the inducing stimulus is strong or weak.* Thus, when the inducing stimulus is more intense, the distance function is steeper. This property will be utilized in the theory below.

Contour Inhibition. Thus far we have described only inhibition that is a function of luminance levels. There is, however, an inhibitory effect which is more properly described as that of one

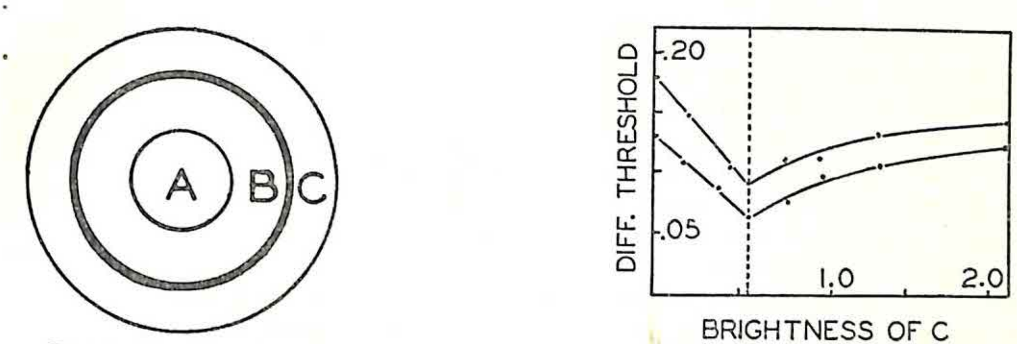

FIG. 8. By either increasing or decreasing the luminance of Annulus C, we can introduce a contour at the B-C junction. (This appears to interfere with the differential brightness threshold at the A-B junction. The lowest threshold is obtained when the B and C luminances are equal, that is, when there is no contour between B and C. The results are shown in the figure on the right—from Dittmers, 1920.)

contour edge on another, and this effect will be of particular interest to us. It has been observed that the resolution of two bright parallel bars is improved by increasing their intensity (Fry & Cobb, 1935). For wide bars (33 minutes of arc) acuity was continuously increased at higher luminance levels. However, with thinner bars (3 minutes of arc) acuity sharply declined with luminance levels above about .5 log trolands. The interprepretation was made that acuity declines in the case of the thin bars, at higher luminance levels, because the two edges of each thin bar, 3 minutes of arc apart, were inhibiting one another. The effectiveness of this variable was confirmed by Fry and Bartley (1935) using a test situation depicted in Figure 8. The threshold of a difference in brightness between areas B and A is determined. The introduction of a dark ring at the outer border of B with C and concentric with A is found to raise that threshold. The threshold rise is greater the closer the dark ring is to the AB border. It appears that the B-C *edge* is exerting an inhibition on the A-B *edge,* thus making its detection more difficult. A very similar type of contour inhibition seems to play a predominant role in one type of metacontrast (Alpern, 1952; Kolers & Rosner, 1960; Werner, 1935) where one edge interferes with an adjoining one presented somewhat earlier in time (usually on the order of 100 milliseconds). Metacontrast contour interference is more powerful when the two contours are closer (Kolers & Rosner, 1960).

There is an additional difference between simultaneous contrast and contour inhibition which is pertinent. Brightness contrast is an asymmetrical phenomenon: Bright areas inhibit dimmer ones, but dim areas have little effect on brighter ones (Graham & Granit, 1931; Heinemann, 1955). Contour inhibition, however, is much more symmetrical: A white edge on a gray background exerts about as much inhibition on the detection of a dark line as does a black edge on a gray background. In an experiment by Dittmers (1920) a stimulus situation very similar to that in Figure 8 was used. The effect of varying the brightness of C upon the differential threshold for A was investigated. The results are also shown in Figure 8 (right). The threshold is lowest when the luminance of B and C are equal, that is, when no edge is present between B and C. Making C either brighter or darker interferes with the establishment of an edge contour between A and B. This must mean that it is the luminance discontinuity per se at the B-C edge which interferes, partly independent of the direction of the gradient. Fry and Bartley (1935) have obtained the same symmetry with a basically similar stimulus. Since figural aftereffects also show such a symmetry—approximately equal displacement effects are obtained with white-on-black or black-on-white inspection figures—inhibition must have a similar symmetry if it is to be invoked as the mechanism of displacement. Simpler simultaneous contrast does not have this symmetry: White inhibits dim regions, but darker areas do not inhibit the brighter ones (Heinemann, 1955). For example, Graham and Granit (1931) found that two adjoining semidiscs of unequal brightness affected each other's critical flicker fusion (CFF) asymmetrically. The CFF of the less bright areas diminished slightly (about 1.5%) while that of the brighter area increased (about 5.4%).

Physiological Evidence. The physiological analysis of inhibition has paralleled the psychophysical data so clearly, although it trails the latter in precision, that it sems worthwhile to present some of that evidence. For one of the advantages of an inhibitory model of contour displacement is precisely that the mechanism is well documented physiologically, and in some detail. The physiological analysis of inhibitory mechanisms in the visual system begins with the work of Hartline (1940) who measured the receptive fields of single optic fibers of the frog, and therefore of single ganglion cells, using very small spots of light. Small areas of stimulation were shown to follow Ricco's law of areal summation. But increasing the area beyond a certain point actually increased the threshold. This was one of the earliest neurophysiological indications of the operation of a visual inhibitory mechanism. Kuffler (1953) demonstrated the existence of an antagonistic (inhibitory) annulus surrounding the receptive field in the retina of the cat. A small spot of light in the center of a receptive field might give an on response, but if the spot was moved to the periphery, the same cell showed an off response. The two spots together gave a much weaker response, the stronger spot luminance determining whether the response would take on the characteristic of the center or the periphery. Thus, a true antagonistic mechanism was at play. Barlow (1953) has shown, on the frog's retina, a similar mechanism at work. First, the response of a spot of light is measured when placed in the center of the receptive field. Next, a second spot is introduced which, although it registers no response itself, inhibits the response to the centrally placed spot. Furthermore, *this inhibitory action is proportional to the intensity of the second spot.* Because we are interested in the proactive inhibitory effects of a stimulus, it is interesting to cite one of Kuffler's (1953) observations. He noted that the suppression of an off response, by stimulation of an on area of a receptive field, could be seen for some time after the on area's stimulus had been turned off. "It was found that beam (A) to an on area could suppress (B) stimulation of an off area for varying periods after the (A) had been turned off. *The time course of the inhibitory after effect depended on the duration and the intensity of (A)* [p. 53; italics mine]." Conversely, a ganglion cell which gave an off effect was often found not to respond to a small spot of light directed to the on area during the period of the off discharge. Thus, the antagonistic action of a receptive field has been found to outlive the duration of the stimulus.

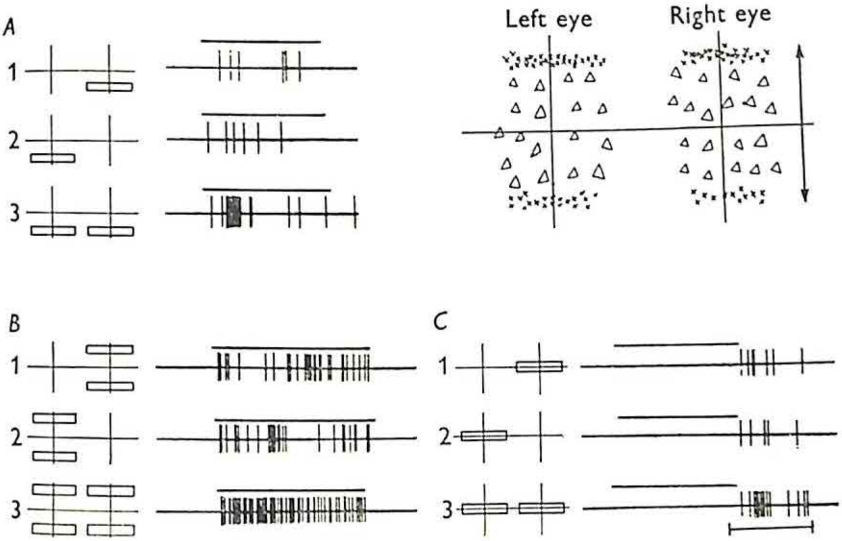

FIG. 9. The vertical lines are action potentials of a single neuron in the visual cortex of the cat. (A horizontal white bar when it first enters the top of the receptive field gives an on response, then an off response at the center of the field, and an on response again at the bottom of the receptive field. This can be obtained from stimulation of either or both eyes, if corresponding retinal points are used. When both eyes are stimulated together, Record 3 in the illustration, an even stronger response is obtained. In Records A and B, the on response is seen to be activated by either eye and to summate binocularly; in C, this is shown to apply to the off response.)

Hubel and Wiesel (1959, 1962) have extended this analysis, examining the receptive fields of cortical neurons. These receptive fields take on complicated shapes and are frequently elongated. Whatever the shape, the cortical neuron gives its strongest response (for so-called "simple" fields) only when numerous light spots are used to summate the activity within the field. (For example, a narrow, slit-shaped stimulus oriented vertically within a vertical receptive field gives a maximal response.) As with retinal cells, and also lateral geniculate cells, these cortical cells have receptive fields with flanks of an inhibitory character. But what makes the response of cortical cells distinct is that line stimuli or edges are the more powerful stimulators, and closely adjoining or parallel lines and edges are the more powerful inhibitors. Diffuse light gives a minimal response. This clearly appears to provide a mechanism for edge inhibition per se, such as has been described above.

Binocular summation is another distinctive property of the cortical cells. Summation occurs between similarly oriented stimuli, such as two horizontal lines presented to corresponding retinal areas of the two eyes; inhibition occurs when a stimulus is presented to the inhibitory flank of the other eye. This is shown in Figure 9, taken from Hubel and Wiesel (1959). The mutual suppression of closely adjoining parallel contours has long been known in psychological literature. (For the binocular suppression of contours, see Asher, 1953; Day, 1961.) But this is the first evidence we have of a neurophysiological mechanism. It provides a clear physiological basis for a theory in which a contour process originating in one eye and a second contour process originating in the other eye are both projected cortically and inhibit each other at the cortical site.

Hartline, Ratliff, and Wagner (1956) and Ratliff and Hartline (1959) have made a quantitative analysis of lateral inhibition among single ommatidia of the limulus eye. Although there are many apparent differences between arthropod and vertebrate visual systems, contrast-enhancing mechanisms seem to be present in both (including the Mach bands), and this suggests that we can gain something from a comparison of the two systems, proceeding with due caution.

In the limulus eye, interaction among ommatidia is inhibitory; there is no neural spatial summation. If we measure the response of a single ommatidium and then illuminate neighboring ommatidia, the threshold of the receptor unit being measured is raised, the number of impulses it discharges in response to a suprathreshold flash is diminished, and the frequency with which it discharges impulses during steady illumination is reduced. The following factors affect the magnitude of inhibition:

1. Intensity—the more intense the inhibitory illumination, the deeper and longer is the initial depression in frequency of discharge. *The steady level is linearly related to log intensity of inhibition.*

2. Area—the larger the size of the neighboring area that is illuminated, the greater the slowing down in rate.

3. Distance—the effectiveness of the inhibition diminishes with distance. If inducing and test stimulus are further apart on the surface of the compound eye of the limulus, then the slope relating the intensity of the inducing figure to the amount of inhibition of the test figure is flattened. This also must mean that the slope of inhibition (I) as a function of distance between I and T is steeper when I is more intense.

The effect of induction and test is mutual. The inducing figure's inhibition of the test figure depends on the illumination of the test figure, since the test figure in turn inhibits the inducing figure. Thus, the rate of discharge, r_p, in the receptor "p," set among a population of n interacting receptors, is given by the set of n simultaneous equations of which one is:

$$r_p = e_p - \sum_{j=1}^{n} k_{pj}(r_j - r^o_{pj})$$

$$1 \leqslant p \leqslant n, \quad [1]$$

where e_p is the discharge from receptor p when it is illuminated alone, r_j is the discharge from a neighboring receptor, j ($j = 1, 2$, etc.), *after* interaction with all the other n receptors in its neighborhood, k_{pj} is the coefficient of inhibition of receptor j on receptor p, and r^o_{pj} is the threshold amount of discharge in j just needed to inhibit I. This equation has been found to fit the limulus data with great precision and has successfully accounted for additional phenomena such as disinhibition and the Mach-band phenomenon. For our purposes the following properties of the equation should be noted. Inhibition is a quantity which is simply subtracted from single-unit discharge; each receptor after being itself inhibited contributes its own inhibition in a simple, additive way. Inhibition is a linear function of receptor discharge, and this is a linear function of log intensity (Hartline, 1940). As distances increase, it is empirically found that the proportionality constant k_{pj} is smaller. Thus, with a smaller distance, the function of inhibition and intensity has a steeper slope. Also, at high intensity, inhibition as a function of distance is steeper. This, in turn, must mean that inhibition as a function of distance has a negative slope and, moreover, is not linear, but is exponential or possibly a polynomial of higher order.

To summarize the neurophysiological contribution, separate inhibitory mechanisms have been discovered which seem to underlie both brightness contrast and contour inhibition (including binocular effects). Both types of inhibition diminish with distance. Inhibition is proportional to the log intensity of the inducing stimulus, and the spatial function is steeper for more intense inducing stimuli. Inhibition can sometimes operate as an aftereffect. And, at least in limulus, *the effect is additive.* A certain amount of discharge frequency, a result of neighboring activity, is subtracted, and the resultant is expressed by a simple sum of simultaneous equations.

An Inhibitory Mechanism of Contour Displacement

We suppose first that the neural correlate of a contour can be represented by a ridge of activity with a cross section which, for illustration, is depicted by T in Figure 10. We will assume that the mean of the distribution will determine the apparent position of the contour as it is measured in an alignment task. The variance of the distribution will determine part of the variance of the alignment judgments. We suppose further that I, the inducing figure, exerts inhibition. The effect of inhibition is to subtract excitation from the test figure's distribution. In other words, the apparent position of the test figure in the presence of some other figure, I, is given simply by the net distribution. Net T is shown by the dashed line in Figure 10.

Köhler and Wallach (1944) found equivalent repulsion displacements occurred whether white-on-black or black-on-white inspection figures were used. Although attempts have been made to demonstrate that these are not equivalent (Ganz & Day, in press), the evidence is equivocal. Assuming, in the absence of clear evidence to the contrary, that they are equivalent, it

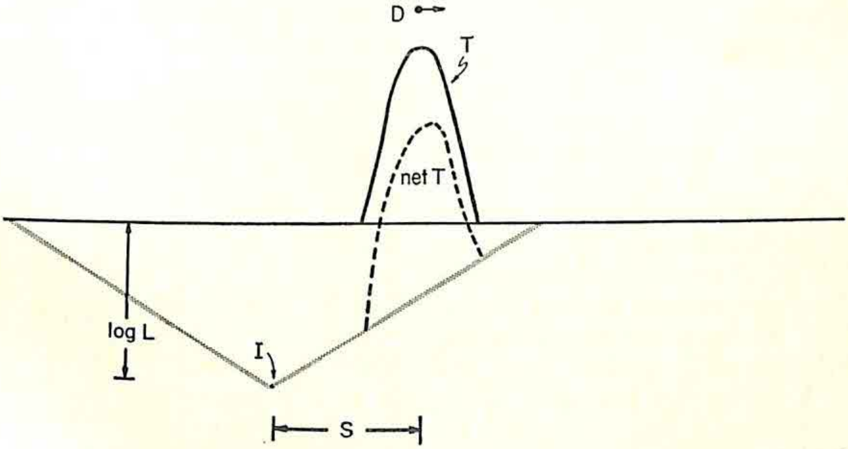

FIG. 10. T represents a theoretical spatial distribution of a cross section of neural excitatory ridge to which a test-figure contour gives rise in the visual system. (The edge of an inducing figure is placed by *E* at a distance, *S*. This I figure is thought to generate lateral inhibition equally on both sides. The inhibition diminishes rectilinearly with distance. The amount of inhibition is proportional to the log contrast of the inducing figure, log L. In this model, the inhibition is subtracted at each point from the T distribution to give a residual excitation, net T. Displacement, *D*, occurs because the mean of the net-T distribution is shifted away from T.)

appears that the type of inhibition which is exerted by I on T which causes it to be displaced is the type restricted to the action of contours on contours, which has been described above.

As a first approximation, it seems plausible (Leibowitz et al., 1953; Ratliff & Hartline, 1959) to assume that for a white line on a grey surround the amount of inhibition is a linear function of the log luminance (log L) of the inhibiting contour. This contour inhibition appears to decay exponentially over a distance of some 120 minutes of arc, with the greatest drop occurring in the first few minutes of arc. For simplicity, we explore a linear model in which the inhibition simply drops linearly to zero at a point m away on both sides of the inhibiting edge. This is shown in Figure 10. If the middle of the test figure is at the center of a coordinate system and the inspection figure at $-S$ on the abcissa, then it can be shown that the amount of inhibition, I, at a point, x, is given by:

$$I_x = 1 - \log \mathrm{L}\left(1 - \frac{S}{m}\right) + \frac{\log \mathrm{L}}{m}x, \quad [2]$$

where L is the luminance of the inducing figure, S is the distance between the inducing and test figures, and m is one-half the width of the total region over which inhibition is exerted by the inducing figure. Log L/m gives the slope of the inhibitory function.

First we would like to show that the actual shape of the test-figure distribution is not a critical variable. In Figure 10 we illustrate the cross section of a straight contour, T, which has a width $-i$ to $+i$. Let the function which relates this cross section to the variable x on the abcissa be $f(x)$. In a control test in which the inducing figure is not presented, the mean of the test-figure distribution is given by:

$$\bar{x} = \int_{-i}^{i} x f(x) dx.$$

Then the net distribution will be:

$$\text{net } \bar{x} = \int_{-i}^{i} x[f(x) - I_x]dx.$$

Substituting Equation 2 into this expression, integrating, and collecting terms, and letting $b = \log \mathrm{L}/m$, we get:

$$\text{net } \bar{x} = \bar{x} - \tfrac{2}{3}bi^3.$$

By definition, contour displacement, D, is given by the mean of the net distribution less the mean of the control distribution, hence:

$$D = -\tfrac{2}{3}bi^3. \quad [3]$$

Since $f(x)$ is not in the equation, it can be seen that the displacement is independent of the particulars of the test figure's distribution, although it is sensitive to the width of the test figure. Moreover, the amount of displacement, D, is proportional to the slope of the inhibition function, b. And since b equals log L/m, we see that contour displacement is (a) proportional to the log of the inducing contour luminance, and (b) inversely proportional to the spread of the inhibition, m.

In the treatment below, we use the convention that the inducing figure is placed at the origin and that the test figure is placed a distance, S, away. If that is the case, and using a test figure of unit width, we get the following distribution:

Test-figure position	*Displacement*
$S \leqslant -m$	0
$-m < S \leqslant 0$	$-b$
$0 \leqslant S < m$	b
$m \leqslant S$	0

Remembering that the inhibition's slope is negative on the left of the inducing figure edge and positive on the right side of the inducing edge, this means that a repulsion displacement occurs on both sides of the inducing figure, and that it is always equal to the slope of the inhibitory function of the same strength, if only the test figure is presented within the region from $-m$ to $+m$.

Role of Eye Movements

A subject's eyes move even though a fixation point is provided, and he makes his best effort to fixate steadily (Ratliff & Riggs, 1950). The subject's eyes move during the inspection-to-test-figure interval, drifting away in all directions from the subject's own mean fixation position. These eye movements have the effect of changing the true interfigural distance at any moment. This must mean that when the results of a block of trials are combined for any one interfigural distance value, set by *E*, the results actually represent the pooling of a whole distribution of interfigural distance values. This distribution will be very broad if the subject drifts a lot from the fixation point and will be quite narrow if the subject has more precise fixation. What then can we expect the subject's eye-movement drifts to be?

Barlow (1952) has taken a measure of the error involved in *O*'s successive fixational pauses. The scatter of eye-movement positions when the subject indicated that he was satisfied with his fixation was 6.5 minutes of arc (5.2 minutes of arc after correction for changes in head position). Two other direct-measurement techniques yielded values ranging from 6.5–5.7 minutes before correction. A subjective method, in which *O* estimates the vertical distance between his fixation point and the afterimage of a previous fixation point, yielded a standard-deviation value of 6.3 minutes of arc.

We are now in a position to examine the figural aftereffect situation. Let σ be the standard deviation of the subject's fixation position in the horizontal plane. From Barlow's data, σ equals about 5 or 6 minutes of arc. We consider the situation in which *E* sets the average interfigural distance at some value less than σ. If the inducing figure has been presented left of the test figure, then, during postinspection, the test figure will be displaced an amount b to the *right* (repulsed) whenever the eye movements bring the test figure momentarily right of the edge of the afterimage of the inducing figure (but not more than m away from I). A definite probability can be attached to the event that an eye movement momentarily brings the test figure left of the afterimage edge (but not more than m away from that edge), and then the displacement will be a repulsion an amount b, but in a leftward direction. Thus, the momentary displacement is not determined, as in previous theories, but is probabilistic, depending on the errorlike vagaries of fixation scatter. For an alignment response, which normally requires multiple fixations, the expected displacement would be the sum of the products of displacements at various *t*-figure positions each multiplied by the probability of the *t* figure assuming such a position. Assuming the distribution of fixation positions to be normal, with a standard deviation defined as σ, we get as the equation of expected displacement[2], D_S:

---

[2] For computation, it is convenient to use the following transformation:

$$D_S = \frac{b}{\pi}\left[2\int_0^{\frac{S}{\sqrt{2\sigma}}} e^{-\gamma^2}d\gamma + i\int_0^{\frac{m-S}{\sqrt{2\sigma}}} e^{-\gamma^2}d\gamma - \int_0^{\frac{m+S}{\sqrt{2\sigma}}} e^{-\gamma^2}d\gamma\right],$$

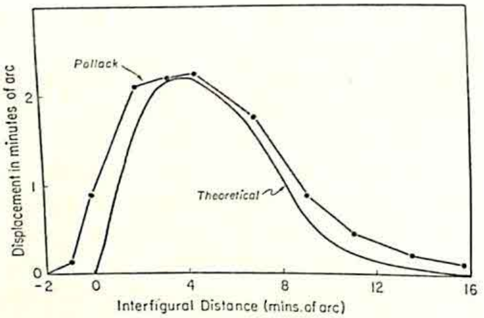

FIG. 11. A comparison of the distribution of figural aftereffects, as they are predicted by the inhibitory model ($\sigma = 5$; $m = 4$) with the results of an experiment (Pollack, 1958).

$$D_S = \frac{b}{\sigma\sqrt{2\pi}}\left[\int_o^m e^{-\frac{(x-S)^2}{2\sigma^2}}dx - \int_{-m}^{o} e^{-\frac{(x-S)^2}{2\sigma^2}}dx\right]. \quad [4]$$

We have plotted the expected displacement when $m = 4$ minutes of arc, and $\sigma = 5$ minutes of arc (after Barlow's results) in Figure 11. It compares favorably with Pollack's (1958) data on figural aftereffects which are also shown in the same figure. It can be seen that the course of the displacements as we vary the interfigural distance is predicted to a first approximation by the theory. The prediction is made on the basis of only one completely free parameter, the constant L which relates the contrast of the inducing figure to its inhibitory potential, the parameter m being undetermined but bounded by empirical data, and the parameter σ being completely determined, and

$$b = \frac{\log L}{m}.$$

Furthermore, the free parameter L is a vertical scale factor and does not otherwise affect the curve.

It might be of interest to note, with regard to the parameter m, that Mach bands, the band of black near a border with white and a band of white near a border with black, have their maximum at about 2–6 minutes of arc (Fiorentini & Radici, 1958) which is about the same location as the point of maximum displacement. The Mach bands are almost certainly an inhibition-disinhibition phenomenon (Ratliff & Hartline, 1959). A recent theory by von Békésy (1960) of the Mach bands, from a schematized summation-inhibition neural unit, used a value of r (roughly equivalent to m in our theory) of 10 minutes of arc. The Leibowitz et al. (1953) data would indicate that the greater proportion of inhibition is dissipated in the first 10 minutes of arc. The discrepancy between the suggested values of 10 minutes of arc from the Leibowitz et al. and von Békésy studies and the 4 minutes of arc which is optimal in the present model probably means that a linear model is to that extent inadequate. Although an exponential decay function of inhibition fits the empirical data on inhibition and figural aftereffect better, the linear model has the advantage of simplicity. However, in the absence of more extensive data on the displacement-distance function, more precision does not seem warranted at present. Moreover, it is our primary interest to show how much of the variance can be accounted for by the simplest model.

Effect of an Inducing Figure on the Variance of Alignment Judgments

The inhibition model can generate an independent and new prediction which is at variance with the expectations based on prevalent theories of

---

where $i = 1$ if $o < S < m$, and $i = -1$ if $m < S$. The bracketed functions are the integrals of the error function:

$$\int_o^x e^{-\gamma^2} d\gamma,$$

which has been tabled (United States National Bureau of Standards Computation Laboratory, 1954).

aftereffects. The inhibition model specified that eye movements sometimes place the retinal projection of a test-figure contour on one side of the inducing-figure contour, sometimes on the other. This will occur a significant proportion of the time when the interfigural distance is less than two or three times the standard deviation of the eye-movement distribution, σ. The displacement of the test-figure contour occurs in opposite direction on the two sides of the inducing-figure contour. Thus, the apparent positions of the test line will be more variable when the interfigural distance is small. This is because, at a small interfigural distance, the test figure is successively placed on both sides of the inducing-figure contour and thus becomes prone to displacements in opposite directions. This in turn will give rise to a bimodal distribution of apparent contour positions, a distribution that has a higher variance than a unimodal distribution. At large interfigural distances, the test figure will cross the inducing-figure boundary only rarely. Thus, it will be exposed to displacements predominantly in one direction. This, in turn, will produce a unimodal distribution which would result in a smaller variance of alignment judgments.

We can derive an explicit expression for the variance of *O*'s contour alignments solely attributable to the effects of the inducing figure. In a control run when no inducing figure is shown, *O*'s alignments are some direct function of the mean of the excitation distribution which the test figure arouses, $\bar{x}$. If *E* places an inducing figure at a particular interfigural distance, *S*, then the expected displacement would be D_S (see Equation 4). Thus, *O*'s *average* alignment would be a direct function of $\bar{x} + D_S$. If the test figure is presented briefly (.1 second or less; a brief enough presentation so that eye movements would not cause significant displacements during the time interval) and happens to fall right of the inducing figure's afterimage (but not more than m minutes away), it would be displaced an amount b. Then the apparent position on that trial would be given by $\bar{x} + b$. Thus, the deviation from *O*'s average judgment, attributable solely to the displacement, would on that trial be given by $(\bar{x} + b) - (\bar{x} + D_S)$, or $(b - D_S)$. Depending on where the test figure falls, during that brief presentation, we have the following deviations:

Test-figure position	*Deviation*
$x < -m$	$0 - D_S$
$-m \leqslant x < 0$	$-b - D_S$
$0 < x \leqslant m$	$b - D_S$
$m < x$	$0 - D_S$

The expected variance, for a specific interfigural distance, *S*, is then given by the product of each of these deviations squared and their probabilities:

$$V_S = (b + D_S)^2 \frac{1}{\sigma\sqrt{2\pi}} \int_{-m}^{o} e^{-\frac{(x-S)^2}{2\sigma^2}} dx + (b - D_S)^2 \frac{1}{\sigma\sqrt{2\pi}} \int_{o}^{m} e^{-\frac{(x-S)^2}{2\sigma^2}} dx. \quad [5]$$

If we let

$$P_a{}^b = \int_a^b e^{-\frac{(x-S)^2}{2\sigma^2}} dx,$$

then Equation 4 becomes

$$D_S = b(P_o{}^m - P_m{}^o).$$

Substituting into Equation 5 and collecting terms, we get

$$V_S = (b^2 + D_S{}^2) P_m{}^m. \quad [6]$$

Substituting the constants we used above to compute D_S into Equation 6, we obtain a function which is plotted in Figure 12. The figure shows a

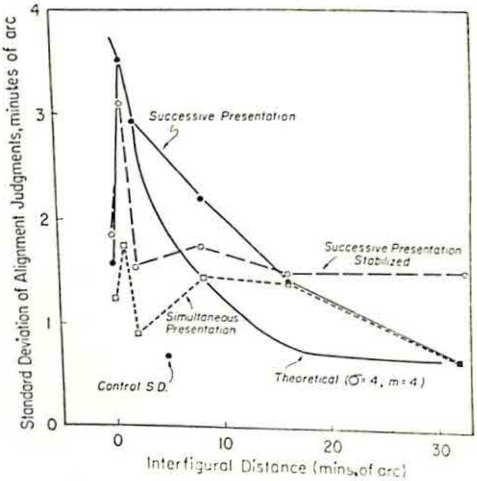

FIG. 12. The standard deviation of vernier alignment judgments (control value) and of vernier alignments when one of the lines is in the vicinity of an inducing figure. (As the distance separating inducing from test figure is made smaller, the standard deviation is seen to increase, especially in the successive presentation. Predicted values, derived from the inhibitory model, using the same parameter values as in Figure 11, are also shown.)

thoretical function which peaks somewhere at a very small interfigural distance, the value at an interfigural distance of 0 being indeterminate, and then drops exponentially reaching a value which is approximately equal to 0 at about 30 minutes of arc. There are, of course, other sources of variance in *O*'s judgments. These other sources can be estimated from alignment judgments performed in the absence of inducing figures. The different variance sources will then simply add in the figural aftereffect situation.

Since we are interested in the variability due to the scatter of momentary test-figure placements, we want to compare our theoretical function to data obtained in experiments in which the test figure has been presented very briefly. If the test figure is presented for a longer time, then *O* can sample a number of eye-movement positions, and thus the variability of test-figure positions due to eye movements will be reduced as the square root of the number of samples. This follows since a judgment would then be based effectively on the mean number of positions. Thus, the mean would represent the standard deviation of a sample mean, that is, the standard error. The results of an experiment (Ganz, 1964) in which the test figure was presented for .12 second is shown in Figure 12. The descending character of the standard deviation in the successive presentations as *S* is increased is quite evident. A second study by Ganz and Day (in press) utilized a method of limits in which each stimulus presentation lasted .3 second. The results are shown in Figure 13. The standard deviation again is seen to fall as *S* is increased. In absolute level, the standard deviation is much smaller, but if we remember that each exposure was about 3 times as long and that each threshold was determined by 4 runs (ABBA in order), then it would appear that we should multiply these values by the square root of 12 to obtain a comparable estimate of the

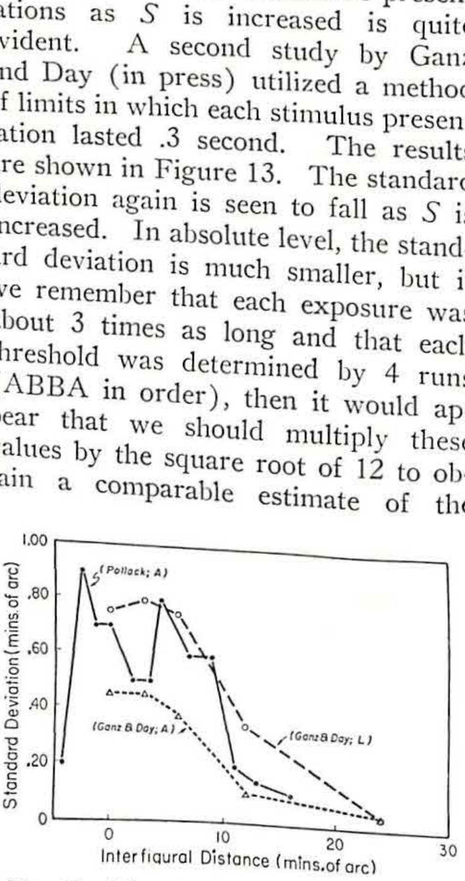

FIG. 13. The standard deviation of vernier alignment judgments at various distances from the location of a previously presented inducing figure. (A signifies vernier judgments obtained with a method of average error where the test figure is presented for an undetermined duration; L signifies the use of a method of limits, with a .3-second test-figure exposure. Compare with Figure 12, where shorter exposures were used.)

standard deviation on single brief exposures. When this is done, the height of the theoretical curve and of the curve from the study by Ganz and Day become quite comparable (the results depicted have *not* been so adjusted). Finally, the data of Pollack (1958) and of Ganz and Day (in press), both studies using a method of adjustments, are shown in the same figure. Here the method of adjustment was used, and in these cases it is not possible to estimate the duration of the test-figure exposure period. Notice, however, that both studies yielded descending functions as the parameter S increased. Also, both functions are slightly lower in magnitude than the function obtained with the method of limits. This is understandable when it is realized that the method of adjustment probably gives the subject the longest stimulus exposure of all these conditions.

It is felt that the prediction of increased standard deviation of alignment near the inducing-figure contour border is specific to an inhibitory model such as the one presented here. It does not follow from the other theories of the figural aftereffect (Köhler & Wallach, 1944; Osgood & Heyer, 1955) where displacements are held to be very small near the inducing-figure border. Thus, the oscillations of the test figure on both sides of the contour of the inducing figure when the interfigural distance is very small would not lead to greater variations in alignment. These theories predict greater variability of alignment where the slope of the displacement function is steepest. This occurs at least at two points, each the same distance from the inducing figure's edge on either side, and at the point of maximum displacement. Put differently, at 0 interfigural distance, the standard deviation of the displacement should be about the same as at a very large interfigural distance since in both cases there is very little displacement. But as Figures 12 and 13 show, this is in fact not the case. Only a theory in which there is strong displacement near the interfigural distance of zero, as in the present model, could predict the type of descending standard-deviation function we have found.

Notice that our equation for D_S implies that the paradoxial distance effect is not inherent in the displacement phenomenon. It evolves from intrinsic errors that arise from the oscillation in the actual momentary test-figure position. These are caused by eye movements and other errors of fixation, or even possibly from errors in E's positioning of the test figure from trial to trial. The latter seems a reasonable possibility since 1 or 2 minutes of arc are frequently close to the limits of error of the apparatus employed (Ganz, 1964; Köhler & Wallach, 1944; Pollack, 1958).

In Taylor's (1962) theory the displacements occur because the subject's alignment judgments are *better* in the presence of the inducing figure. Taylor feels that the inducing figure acts as an anchor, and this effect is stronger when the interfigural distance is small. It is unlikely, however, that more accurate judgments should come to be associated with greater variance.

An observation was often made, especially under conditions of very careful fixation (Ganz, 1964), that during the postexposure period the apparent position of the test figure changed *while O observed it,* often from one side of the inducing figure's edge to the other side. This is not accounted for except in a model that includes the effects of sudden corrective eye-movement saccades which might bring the test figure to the other side of the inspection contour during the observation period.

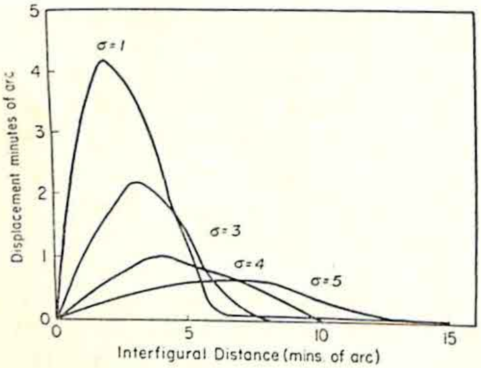

FIG. 14. The effect of varying the scatter of horizontal eye movements (σ) on the distribution of figural aftereffects, as derived from the inhibitory model.

If our theory is correct, the distribution of displacements should change if the scatter of eye movements is altered. One experimental way of accomplishing this is to stabilize the retinal image. The result of such stabilization should be to restrict the variance increases of the alignment judgments largely to the region in the immediate vicinity of the inducing-figure edge. This would be the result predicted from Equation 6 if the parameter σ is reduced in value. This is shown in Figure 14. An experiment on image stabilization (Ganz, 1964) confirmed this expectation. The results of this experiment are given in Figure 12: successive presentation, stabilized.

We should also consider the case where, although eye movements are not reduced, the test figure does not vary in its position with respect to the inducing figure. This would be the case when the inducing and test figure are presented together. In such a situation, the increase in alignment variability at small interfigural distances should be absent. In Figure 12 the data obtained under the simultaneous presentation show only very small variance increments. To this extent our expectations appear to be confirmed.

Figural Aftereffects as Simultaneous Illusions

The theoretical model presented here considers the displacements which occur in the figural aftereffect situation to be a consequence of a simultaneous illusion—mutual contour inhibition—brought about by the decaying trace of the afterimage upon the test contour. We have shown that a theoretical model of mutual contour inhibition does predict displacements of appropriate location and variability. One would like to know whether such displacements occur when an actual inducing-stimulus object is substituted for the postulated afterimage. In other words, do simultaneous illusions occur which are similar in direction and magnitude to the figural aftereffect?

Presenting an inducing and test figure together (unlike the afterimage situation) effectively reduces the magnitude of the eye-movement component. During the simultaneous illusion, any eye movement would move both inducing and test figure together across the retina. Thus, the inducing figure would retain its distance and direction with respect to the test figure. It is unlikely that the effective eye movements would be completely nullified, since contour inhibition appears to act with a definite latency (Kolers & Rosner, 1960). In Figure 14 we have depicted the figural aftereffect distributions which the theory predicts when the parameter σ, the standard deviation of the eye position during fixation, is reduced. We see that as the peak of the point of maximal displacement approaches the edge of the inducing figure the peak is steeper, and its magnitude is greater. In Figure 4 we see the results of an experiment with 30 subjects in which the inducing and test figures were presented simultaneously

(Ganz, 1964). Two successive series of measurements are shown. Pollack's (1958) experiment in which similar figures were presented in the aftereffect paradigm is also included in the same figure for comparison. We see that the simultaneous experiment did yield a point of maximum displacement nearer to the inducing-figure edge (1 minute of arc) than the value obtained from Pollack's data (about 4.5 minutes of arc) or to any figural aftereffect experiment known to the author.

It should be noted, finally, that the Inhibition Model predicts that contour displacements will occur with or without fixation tremor. It is only the magnitude of the displacements and the point of maximal displacement which change with such errors of fixation. Thus, the theory is not inconsistent with the fact that figural aftereffects are found under conditions of retinal image stabilization (Ganz, 1964; Hochberg & Hay, 1956; Krauskopf, 1960).

SUMMARY

The edges of objects in the visual field appear to repulse one another when they are brought close together. These mutual shifts have long been known to us in the case of the classical geometrical illusions. Such effects can be obtained even when a pair of neighboring contours is presented successively. These are known as the figural aftereffects.

A quantitative theory is proposed to account for these perceptual aberrations. The two aspects of the aftereffect—the shifts in contour position and the temporal separation which can be introduced between inducing and test contour—are treated separately. According to the theory, the displacement arises because the neural correlates of visual contour inhibit one another. The inhibition is proportional to the log of the contour intensity (more exactly to its contrast above or below background level). The inhibition is a decreasing linear function of the separation between the contours. These two assumptions, both documented psychologically and physiologically, are sufficient to generate a predicted contour displacement. Eye movements during fixation, also documented, are shown to produce a statistical distribution of repulsions at various intercontour distances. By introducing empirically derived values for the equation parameters, a distribution of figural aftereffects is generated which agrees well with experimental values.

Two independent sets of predictions are generated by the model. First, the model predicts a very steep rise in response variability as the test contour is brought closer to the inducing contour (particularly when the two are presented successively). Second, certain changes in the mode of presentation—retinally stabilized, simultaneous presentation, short test-figure presentation—are predicted to yield different distributions of figural aftereffects. The data available to date corroborate the predictions.

The mechanism which bridges the time gap is described only briefly. We hope to consider it in more detail in a subsequent communication. The figural aftereffects are, it is believed, simultaneous illusions which are in many ways no different from the classical geometrical illusions. When the inducing figure is removed, an afterimage of the inducing figure remains in *O*'s visual field. The neural correlates of that afterimage then repulse the test figure's neural correlate, in accordance with the displacement mechanism just described. The lines of evidence which support an after-

image explanation are strongly suggestive, though perhaps not conclusive. That is, it is found that the onset, the initial intensity, and the decay of figural aftereffects and afterimages are strikingly similar, often involving the same parameter values. They, in turn, both appear to reflect the operation of light- and dark-adaptation mechanisms.

Thus, it appears possible that most of the quantitative properties of the figural aftereffects will be reduced to the operaton of a number of basic visual mechanisms: adaptation, inhibition, and ocular tremor.

REFERENCES

ALPERN, M. Metacontrast: Historical introduction. *American Journal of Optometry,* 1952, **29,** 631–646.

ASHER, H. The suppression theory. *British Orthoptic Journal,* 1953, **10,** 1–9.

BARLOW, H. B. Eye movements during fixation. *Journal of Physiology,* 1952, **116,** 290–306.

BARLOW, H. B. Summation and inhibition in the frog's retina. *Journal of Physiology,* 1953, **119,** 69–88.

BARLOW, H. B., & SPARROCK, J. M. B. The role of after-images in dark adaptation. *Science,* 1964, **144,** 1309–1314.

BARTLEY, S. H. The psychology of vision. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 924–926.

BEITEL, R. J. Inhibition of threshold excitation in the human eye. *Journal of General Psychology,* 1936, **14,** 31–61.

BÉKÉSY, G. VON. Neural inhibitory units of the eye and skin: Quantitative description of contrast phenomena. *Journal of the Optical Society of America,* 1960, **50,** 1060–1070.

BORING, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century, 1942.

BOYNTON, R. M., & RIGGS, L. A. The effect of stimulus area and intensity upon the human retinal response. *Journal of Experimental Psychology,* 1951, **42,** 217–226.

BRINDLEY, G. S. After-images. *Scientific American,* 1963, **209,** 84–93.

DAY, R. H. On the stereoscopic observation of geometrical illusions. *Perceptual and Motor Skills,* 1961, **13,** 247–258.

DAY, R. H. Excitory and inhibitory processes as the basis of contour shift and negative after-effect. *Psychologia,* 1962, **5,** 185–193.

DAY, R. H., POLLACK, R. H., & SEAGRIM, G. N. Figural after-effects: A critical review. *Australian Journal of Psychology,* 1959, **11,** 15–45.

DEUTSCH, J. A. The statistical theory of figural after-effects and acuity. *British Journal of Psychology,* 1956, **47,** 208–215.

DEUTSCH, J. A. Neurophysiological contrast phenomena and figural after-effects. *Psychological Review,* 1964, **71,** 19–26.

DITTMERS, F. Über die abhängigkeit der unterschiedesschwelle für Helligkeiten von der antagonistischen Induktion. *Zeitschrift für Sinnesphysiologie,* 1920, **51,** 214–232.

FIORENTINI, ADRIANA, & RADICI, TILDE. Brightness, width and position of Mach bands as a function of the rate of variation of the luminance gradient. *Atti della Fondazione Giorgio Ronchi,* 1958, 13, 145–155.

FLAMANT, FRANCOISE. Étude de la repartition de lumiere dans l'image retinienne d'une fente. *Revue Optique,* 1955, **34,** 433–459.

FRY, G. A., & BARTLEY, S. H. The effect of one border in the visual field upon the threshold of another. *American Journal of Physiology,* 1935, **112,** 414–421.

FRY, G. A., & COBB, P. W. A new method for determining the blurredness of the retinal image. *Transactions of the American Academy of Ophthalmology and Otolaryngology,* 1935, **40,** 423–428.

GANZ, L. Lateral inhibition and the location of visual contours: An analysis of figural after-effects. *Visual Research,* 1964, **4,** 465–481.

GANZ, L., & DAY, R. H. An analysis of the satiation-adaptation mechanism of figural after-effects. *American Journal of Psychology,* 1966, in press.

GRAHAM, C. H., & GRANIT, R. Comparative studies on the peripheral and central retina: VI. Inhibition, summation, and synchronization of impulses in the retina. *American Journal of Physiology,* 1931, **98,** 664–673.

HARTLINE, H. K. The effects of spatial summation in the retina on the excitation of the fibers of the optic nerve. *American Journal of Physiology,* 1940, **130,** 700–711.

HARTLINE, H. K., RATLIFF, F., & WAGNER, H. G. Inhibition in the eye of the limulus. *Journal of General Physiology,* 1956, **39,** 651–673.

Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.

Heinemann, E. G. Simultaneous brightness induction as a function of inducing and test-field luminances. *Journal of Experimental Psychology,* 1955, **50,** 89–96.

Hess, C., & Pretori, H. Messende Untersuchungen über die Gesetzmässigkeit der simultanen Helligkeitcontrastes. *Albrecht von Graefe's Archiv für Ophthalmologie,* 1894, **40,** 1–24.

Hochberg, J. E., & Hay, J. Figural aftereffect, after-image, and physiological nystagmus. *American Journal of Psychology,* 1956, **69,** 480–482.

Hubel, D. H., & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology,* 1959, **148,** 574–591.

Hubel, D. H., & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology,* 1962, **160,** 106–154.

Ikeda, H., & Obonai, T. The quantitative analysis of figural after-effect: I. The process of growth and decay of figural after-effect. *Japanese Journal of Psychology,* 1953, **24,** 59–68.

Ikeda, H., & Obonai, T. Studies in figural after-effects: IV. The contrast-confluence illusion of concentric circles and the figural after-effect. *Japanese Psychological Research,* 1955, **2,** 17–23.

Ikuta, H. Displacement in figural aftereffect and simultaneous illusions. *Japanese Journal of Psychology,* 1956, **27,** 218–226.

Köhler, W., & Wallach, H. Figural after-effects: An investigation of visual processes. *Proceedings of the American Philosophical Society,* 1944, **88,** 269–357.

Kolers, P. A., & Rosner, B. S. On visual masking (metacontrast): Dichoptic observation. *American Journal of Psychology,* 1962, **73,** 2–21.

Krauskopf, J. Figural after-effects with a stabilized retinal image. *American Journal of Psychology,* 1960, **73,** 294–297.

Kuffler, S. W. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysioloy,* 1953, **16,** 37–68.

Lashley, K. S., Chow, K. L., & Semmes, J. An examination of the electrical field theory of cerebral integration. *Psychological Review,* 1951, **58,** 123–136.

Lehmann, A. Über die Anwendung der Methode der mittleren Abstufungen auf den Lichtsinn. *Philosophische Studien,* 1886, **3,** 497–533.

Leibowitz, H., Mote, F. A., & Thurlow, W. R. Simultaneous contrast as a function of separation between test and inducing fields. *Journal of Experimental Psychology,* 1953, **46,** 453–456.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. What the frog's eye tells the frog's brain. *Proceedings of the Institute of Radio Engineers,* 1959, **47,** 1940–1951.

Logan, J. A. A general approach to the study of visual illusions and figural aftereffects. *Australian Journal of Psychology,* 1960, **12,** 235–236. (Abstract)

McEwen, P. Figural after-effects. *British Journal of Psychology Monograph Supplement,* 1958, No. 31.

Nozawa, S. Prolonged inspection of a figure and the after-effect thereof. *Japanese Journal of Psychology,* 1953, **23,** 217–234. (a)

Nozawa, S. Prolonged inspection of a figure and the after-effect thereof. *Japanese Journal of Psychology,* 1953, **24,** 47–58. (English summary, 84–85) (b)

Osgood, C. E., & Heyer, A. W. A new interpretation of the figural after-effect. *Psychological Review,* 1952, **59,** 98–118.

Oyama, T. Experimental studies of figural after-effects: I. Temporal factors. *Japanese Journal of Psychology,* 1953, **23,** 239–245.

Oyama, T. Temporal and spatial factors in figural after-effects. *Japanese Psychological Research,* 1956, **3,** 25–36.

Pollack, R. H. Figural after-effects: Quantitative studies of displacement. *Australian Journal of Psychology,* 1958, **10,** 269–277.

Ratliff, F., & Hartline, H. K. The responses of *Limulus* optic nerve fibers to patterns of illumination on the receptor mosaic. *Journal of General Physiology,* 1959, **42,** 1241–1255.

Ratliff, F., & Riggs, L. A. Involuntary motions of the eye during monocular fixation. *Journal of Experimental Psychology,* 1950, **40,** 687–701.

Smith, K. "Attraction" in figural aftereffects. *American Journal of Psychology,* 1954, **67,** 174–176.

Sperry, R. W., & Miner, Nancy. Pattern perception following insertion of mica plates into visual cortex. *Journal of Comparative and Physiological Psychology,* 1955, **48,** 463–469.

SPERRY, R. W., MINER, NANCY, & MEYERS, R. E. Visual pattern perception following subpial slicing and tantalum wire implantations in the visual cortex. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 50–58.

SPITZ, H. H. The present status of the Köhler-Wallach theory of satiation. *Psychological Bulletin,* 1958, **55**, 1–28.

TAYLOR, M. M. Figural after-effects: A psychophysical theory of the displacement effect. *Canadian Journal of Psychology,* 1962, **16**, 247–277.

UNITED STATES NATIONAL BUREAU OF STANDARDS, COMPUTATION LARORATORY. Tables of the error function and its derivative. Washington, D. C.: United States Government Printing Office, 1954.

WERNER, H. Studies on contour: I. Qualitative analyses. *American Journal of Psychology,* 1935, **47**, 40–64.

(Received October 8, 1964)

Psychological Review
1966, Vol. 73, No. 2, 151–163

SEXUAL MOTIVATION[1]

RICHARD E. WHALEN

Department of Psychobiology, University of California, Irvine

Current theories of sexual motivation emphasize the hormonal determinants of animal motivation and the experiential determinants of human motivation. The present theory suggests that both human and animal sexual motivation are controlled by both biological and experiential determinants. Sexual motivation is said to be comprised of 2 components, sexual arousal, which represents the current level of sexual excitation, and sexual arousability, the propensity for arousal. The sexual motivation of an individual may be represented by a specification of the state of these 2 components. Arousal is modulated by the presence and absence of relevant stimuli. Arousability is determined by hormonal state, feedback effects of copulation, and experience. The present theory stresses the difference between the state of sexual motivation and the expression of sexual behavior.

Students of human behavior usually find little relationship between that aspect of human sexuality which interests them and the sexual behavior of animals (e.g., Hardy, 1964). It is common, for example, to contrast the immediate cessation of mating in female rodents following ovariectomy with the occasionally reported increase in libido of female humans following removal of the ovaries. It is also common to contrast the congruence between type of gonad and mating patterns in animals with the absence of congruence between type of gonad and gender role displayed by some human hermaphrodites. It is uncommon, however, for students of human behavior to compare analogous facets of human and animal sexuality. Libido is compared with lordosis, the reflex posture assumed by estrous female rats and cats when mounted by the male, even though libido and lordosis represent quite different aspects of sexuality. Libido is not compared with the sexual desire of the female rat as measured, for example, by the latency to mate when the female is placed with a male, or by her willingness to work for, or to accept pain to reach, a male.

The failure to compare analogous facets of human and animal sexuality seems to arise from three facts. First, human sexual motivation is usually inferred from subjective reports, while animal sexual motivation is usually inferred from the reliable observation and measurement of discrete components of behavior.[2] Second, students of human behavior discuss sexuality in motivational terms, such as libido, while students of animal behavior discuss sexuality in response terms, such as intromission latency or frequency. Third, the motivational terms used to describe human sexuality are often ambiguous. Libido is rarely defined in the literature; Hardy (1964) uses the expression "true sexual desire (lust)," yet fails to provide an adequate definition of this expression. It is certainly not clear that libido and lust are equivalent terms and could be measured by the same scales.

---

[1] The ideas expressed here grew out of research supported by grants and fellowships from the National Institutes of Mental Health and Child Health and Human Development. Although these ideas were greatly influenced by my training and association with Frank A. Beach, only I can be held responsible for them. I thank J. A. Deutsch for keeping me out of one or two theoretical cul-de-sacs through his thoughtful reading and criticism of the manuscript.

[2] Some of our inferences are based upon subjective reports of human sexual behavior. Although such reports are often inaccurate, they provide the only data we have which are related to human sexual motivation.

In the past, a similar problem plagued students of animal behavior (cf. Kuehn & Beach, 1963). Discrete and different responses within the mating pattern were not discriminated, and all mating responses were used equally to infer motivation. However, recent research has shown, for example, that the mounting behavior of the male rat contains three components—mounting without intromission, mounting with intromission, and mounting with ejaculation—and that the occurrence of each of these responses has a different effect upon subsequent mating (Beach & Whalen, 1959a, 1959b; Whalen, 1961). The frequency and timing of each of the mounting responses seem to be related to sexual motivation in a different manner.

Attempts have been made to solve some of these definitional problems. Money (1961) attempted to delimit the concept of human sexual desire, which he considers equivalent to libido, when he said, "sexual desire is manifested in erotic imagery, sensation, report and action [p. 241]." Although we might like a more complete description of these terms and statements relating these terms to their measurement, the definition offered by Money is useful since it excludes some aspects of sexuality from the concept of sexual desire. For example, the loss of ejaculatory ability following gonadectomy would not indicate a loss of libido, even though the loss of ejaculatory ability might result in subsequent decline in sexual desire. Libido, ejaculation, gender role, and masturbation are all part of human sexuality, but they are not all part of sexual motivation.

DIMENSIONS OF SEXUALITY

Of primary importance for the study of human sexuality is a discrimination of the components of sexuality. Sexuality is not a unitary dimension which is determined in any simple way by hormones or by experience. In the past, several classificatory schemes have been proposed for the analysis of sexuality. The one which we will adopt is similar to that proposed by Winoker (1963).

1. Sexual Identification: the gender role, male or female, of an individual. Masculinity and femininity are related to gender role, but are not exclusively determined by it.
2. Object Choice: those persons or objects toward which an individual directs his sexual activities, be these intercourse, fondling, or fantasy.
3. Sexual Gratification: the reinforcement, reward, or "pleasure" associated with or caused by sexual activities.
4. Sexual Arousal: the momentary level of sexual excitation.
5. Sexual Arousability: an individual's characteristic rate of approach to orgasm as a result of sexual stimulation. It must be emphasized that sexual arousability is a rate and not a state function.
6. Sexual Activity: sexual behavior exemplified by reported fantasy and desire as well as by observed or reported overt behavior.

In the discussion which follows we will be concerned primarily with sexual arousal and arousability. We feel that these two dimensions of sexuality comprise sexual motivation.

Unfortunately, sexual motivation is like other motivational states—it cannot be measured directly. Sexual arousal and arousability must be inferred from behavior. This fact raises a major problem. Behavior is not simply a function of motivational state; behavior is a function of some interaction between motivation and habit (Wittenborn, 1957). In species such as the rat where habits usually contribute little to sexual behavior (Rabedeau & Whalen, 1959), measures of the behavior may allow relatively direct inferences about motivation. In carnivores and primates, habits greatly influence the expression of sexual behavior (Harlow, 1962; Rosenblatt & Aronson, 1958; Whalen, 1963a). In such cases the validity of our inferences about sexual motivation is reduced. Because of this difficulty we must seek measures of behavior which are either little influenced by habits or experience, or we must study behavior after habits are firmly established. In short, we feel that while behavior reflects motivation, the course of behavior is not part of the motivation. This approach to sexual motivation differs from that of Hardy (1964) who feels that:

> For an expectation to be motivational in character, it must involve not only affective change, but also a course of behavior, symbolic or overt, which is instrumental to attaining the desired affect [p. 16].

SEXUAL MOTIVATION

Arousal and Arousability

The expression "sexual motivation" is commonly thought of in two ways. First, an individual is said to be sexually motivated at some particular time. This transient component of sexual motivation, which we have termed *sexual arousal,* represents the current level of sexual excitement. Arousal may range from no excitement to a maximal level of excitement. Maximal arousal is characterized by the occurrence of orgasm. Arousal may be measured by the probability of response, latency of response, frequency of response per unit time, and so forth.

The second facet of sexual motivation we have termed *sexual arousability.* Arousability is defined as the rate at which an individual approaches maximal sexual arousal. Arousability is determined not only by the amount of increment in arousal which is produced by each individual sexual stimulus, but also by the absolute number of sexual stimuli which simultaneously arouse the individual, and the number and effectiveness of stimuli which are negatively arousing (or sexually inhibiting). It should be clear that a person characterized as highly arousable may not become aroused at some particular time because of the absence of positive stimuli or the presence of negative stimuli. Net arousability may be measured by the number of stimuli which are needed to produce orgasm.

Figure 1 illustrates the effects of applying a single sexual stimulus to individuals of high and low arousability. Each stimulus produces an immediate increment in arousal which reaches a peak and then decays in time. If a second stimulus is applied before the effects of the first stimulus have completely decayed, sexual arousal will

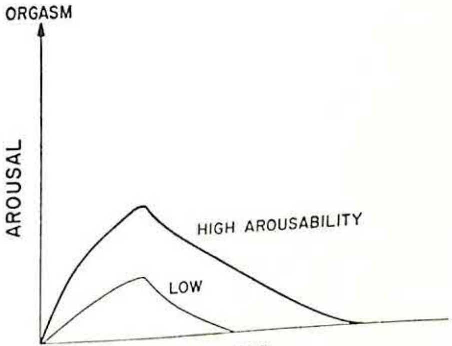

FIG. 1. Increments in and decay of arousal produced in individuals of high and low arousability by a single sexual stimulus.

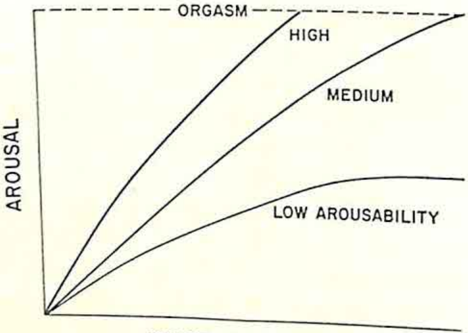

FIG. 2. The growth of arousal in individuals of high, medium, and low arousability produced by a series of consecutive sexual stimuli. (In two instances arousal reaches its maximum level and orgasm occurs.)

progressively increase. Arousal would accumulate more rapidly for the high-arousable individual than for the low-arousable, both because each stimulus would add relatively more excitation and because the decay of arousal would require more time.

Figure 2 illustrates the smoothed growth of arousal as a function of the number of successive stimuli for individuals of high, medium, and low arousability. In this figure the individuals of high and medium arousability would reach orgasm, but at different rates, while the individual of low arousability would approach, but never reach, maximal arousal. For this last individual, each stimulus would add very little to arousal, and since arousal produced by a stimulus would begin to decay immediately, stimuli could not follow one another rapidly enough for the accumulated arousal to reach maximum.

That the effects of successive stimuli do accumulate, do lead to orgasm, and do decay with time has been suggested by work with lower animals (Beach & Whalen, 1959a, 1959b; Bermant, 1964; Larsson, 1956). Beach and Whalen (1959b) showed that for the male rat successive intromissions add approximately equal increments to the excitation needed to produce ejaculation, and that as the delay between one intromission and the next increases, the amount of excitation decreases. If intromissions are spaced far enough apart in time, ejaculation never occurs (Larsson, 1956). These effects of spacing intromissions may be thought to reflect changes in arousal illustrated in Figures 1 and 2. Observed individual differences in the number of intromissions needed to produce ejaculation may be thought to reflect individual differences in sexual arousability.

We are stating that, given a constant habit level, sexual arousal and arousability may be estimated from measures of sexual behavior. It should be clear that our measures of arousal and arousability are not entirely independent. If arousal level is not known, arousability cannot be estimated accurately. For example, a highly aroused individual may reach orgasm with very little sexual stimulation. If one did not have information about the high initial arousal, one might conclude that the individual was simply highly arousable, and reached orgasm with little sexual stimulation. In practice, arousal may be discriminated from arousability, and one must take care to do so.

DETERMINANTS OF AROUSAL AND AROUSABILITY

Arousal

For an individual of any specified level of arousability, increments in sexual arousal are induced by certain external and internal stimuli. A greater number of stimuli can induce arousal in the high- as compared with the low-arousable individual.

We are forced to say only that "certain" stimuli after arousal, for it is impossible to specify the effective stimuli.

Sexual arousal is influenced by stimuli in all sensory modalities—vision, touch, hearing, etc. No one modality is critical for arousal (Ford & Beach, 1951). Furthermore, individuals become conditioned with respect to which stimuli are particularly effective in inducing arousal. Thus, while nearly everyone is responsive to tactile stimulation, only some individuals are responsive to erotic visual stimuli, and some individuals to only some forms of visual stimulation. These multimodality and conditioned individual difference factors prevent predictions about effective arousing stimuli in any particular instance.

Arousability

As with arousal, sexual arousability is determined by learning and by the physiological state of the organism. The physiological determinants of arousability seem to be of two kinds. First are the hormonal determinants of arousability.

The prevalent notion today is that humans differ from other animals in their dependence upon hormones for their sexual behavior. It is accepted that castration has little effect upon human sexual motivation. Particular individuals are said to report either increased or decreased libido following gonadectomy. However, there is evidence which contradicts this notion and shows that in males castration does result in a decline in sexual desire and performance. Beach (1948) after a review of several of the older studies of this question concluded that castration does reduce sexual activity. More recent studies confirm this conclusion. For example, Bremer (1959) completed an extensive study of 215 males who had been legally castrated in Norway. Of the 157 cases on whom there was sufficient information, 74 reported "asexualization" shortly after castration, and 29 others reported "asexualization" within the first year after operation. Only 54 individuals reported *some* sexual interest and activity after the first postoperative year. Bremer concluded:

> It can be stated at the beginning that in all cases without exception the amount of sexual activity has been altered. It has been reduced or abolished irrespective of the (precastrational) direction or the form of the sexual urge . . . [p. 67].

Since many of these individuals continued to be exposed to sexual stimuli yet failed to become aroused, the reported reductions in sexual activity may be assumed to reflect reductions in sexual arousability.

Studies of hypogonadal men under testosterone therapy also support the hypothesis that the male hormones, androgens, exert a major influence upon sexual arousability. Money (1961) after an examination of 11 such cases concluded that: "In each case, the absence of androgen from the tissues made a decided difference to sexual desire and sexual performance [p. 243]." A reduction in the frequency of erotic dreams and imagery as well as a reduction in the frequency of the initiation of intercourse was noted. Castration and hypogonadalism, therefore, reduce the expression of sexual behavior. We assume, tentatively, that these reductions reflect changes in sexual arousability. Additional data are needed for this assumption to be verified.

Of course, one would not wish to conclude, on the basis of these data, that the expression of sexual behavior is solely under the control of androgenic hormones. As noted earlier, experience and habits do influence the expression of sexual behavior. Sexually experienced male cats (Rosenblatt & Aronson, 1958) continue to mate for several months following gonadectomy, and castrated and hypogonadal men continue to have erotic sensations, desires, and fantasies, and do masturbate and occasionally have intercourse (Bremer, 1959; Money, 1961). Habits, in the absence of hormones, can mediate some sexual expression. The point we wish to stress is that in the absence of androgenic hormones the sexual arousability of men, as of other animals, is reduced.

In female organisms the problem of the hormonal contribution to sexual arousability is more complex than in males. The simple assumption that estrogen directly controls arousability may not be true. Sexually naïve female cats primed with estrogen will not accept the male, even though they will assume the mating posture either spontaneously or in response to perineal stimulation (Whalen, 1963a). In this case, the primary effect of estrogen is on the facilitation of the postural reflexes necessary for mating, rather than on arousability per se. This may be an example of the effects of estrogen on mating in general. Estrogen may be crucial for mating only in those species in which unusual postural adjustments are required for successful matting. A dissociation between estrogen secretion and mating seems to occur in those species in which unusual postures are not required for mating. By "unusual" postures we mean those postures, such as lordosis, which have a low probability of occurrence in nonmating situations. Female monkeys and humans accept the male at times other than the time of maximal follicular development and, when mating, assume postures which occur with moderate to high frequency at times other than mating. In these cases, estrogen clearly is not directly related to arousability.

In all species, estrogen may *indirectly* affect arousability through its production of cytological changes in the vaginal epithelium and the concomitant facilitation of vaginal lubrication. Vaginal lubrication may make intercourse less difficult and painful and thereby contribute to arousability. The slight and inconsistent reductions in sexuality which are reported to follow ovariectomy (Bremer, 1959; Fuller & Drezner, 1944) may be accounted for in terms of changes in vaginal state rather than by changes in arousability.

Although sexual arousability may be unrelated to estrogen levels in women, arousability may still be under hormonal control. It has been suggested by a number of investigators that androgen is the primary libido hormone in women (Money, 1961). Waxenberg, Drellich, and Sutherland (1959) found that female cancer patients who underwent therapeutic ovariectomy reported some decrease in sexual desire, activity, and responsiveness suggestive of a reduction in arousability. However, the reported decreases in libido were not nearly as profound as those reported following combined ovariectomy and adrenalectomy. Following the combined operation, 82% reported a decline in desire, 100% a decline in activity, and 92% a decline in responsiveness. The authors suggest that these declines in sexuality were the result of the removal of the androgens which are produced by the adrenal glands. Although these findings are suggestive, one must worry about the validity of reports from women with cancer, castration, and adrenalectomy.

Studies of the effects of administered testosterone may be more valid. Foss (1951) examined 10 women who were being treated with 200–300 milligrams of testosterone daily. Five reported an increase in libido, or sexual craving, two others reported an increased frequency of intercourse, three noted only an increase in the size of the clitoris, and only one individual reported no effect of the treatment. Similar data, obtained by Salmon and Geist (1943) from a study of 101 cases, support the notion that androgen, rather than estrogen, determines sexual arousability in women. We are assuming in these cases that the androgen determines the responsiveness to erotic stimuli. In the absence of androgenic stimulation, arousability would be low, and stimuli would not act to increase arousal.

It should be noted that the locus of the effects of androgen is still unclear. While androgen may be directly influencing arousability by its action on critical neural centers, androgen may also be influencing arousability by its peripheral action. Androgen does cause growth of the clitoris (Foss, 1951; Groome, 1939), and clitoral stimulation does lead to orgasm (Masters, 1960). It is entirely possible that the reported increase in libido following testosterone is the result of increased clitoral sensitivity and the resulting facilitation of orgasm (Carter, Cohen, & Shorr, 1947). Kupperman and Studdiford (1953) concur with this interpretation, but also suggest that the androgen works by increasing general "well being." Money (1961) feels that androgen has both a direct effect upon libido and an effect mediated by increased sensitivity of the clitoris.

Regardless of the mode of action of androgen in women, the data do indicate that hormones affect sexual arousability. The false assumption that estrogen must be the critical hormone underlying arousability may have led investigators (e.g., Hardy, 1964) to the conclusion that hormones are irrelevant to human sexuality.

Unfortunately, data are not available which allow us to make statements about the precise physiological relationships between hormones and arousability. The data which do exist, however, are consistent with the notion that hormones affect arousability by altering the threshold for erotic stimulation, regardless of whether the threshold in question is one of peripheral tissue sensitivity or of central neural sensitivity. It is also reasonable to assume that the hormones act centrally by determining the amount of change in arousal produced by a given stimulus, and act peripherally by determining the amount of receptor response to a stimulus, and/or the rate of adaptation of the receptor. Additional data are needed to verify these points.

A second physiological determinant of arousability, one which is seldom discussed, is the feedback effect of sexual activity upon subsequent arousability. Extensive analyses of rat and cat mating indicate that the performance of sexual responses reduces arousability. This reduction in arousability occurs in two ways. For the male rat, each ejaculation seems to inhibit arousability, and several ejaculations lead to the cessation of mating (Beach & Jordan, 1956b). For the male cat, the feedback effects of ejaculation seem to be minimal. Reductions in arousability do accompany mating in this species, but these reductions seem to result from some process of stimulus adaptation (Whalen, 1963b). In the human male, ejaculation or orgasm seems to affect arousability as it does in the rat. Masters and Johnson (1963) have shown that the human male is resistant to sexual stimulation immediately after ejaculation. Continued postorgasmic stimulation slows, but does not prevent, the detumescence process.

It is important to note that at orgasm and the cessation of stimulation, arousal and arousability do not decay at the same rate. It is likely that, *for the*

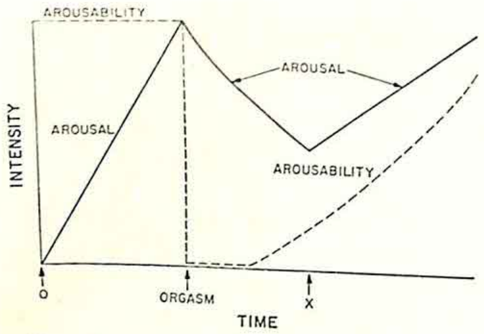

FIG. 3. Changes in arousal and arousability as a function of sexual stimulation and orgasm. (See text for full description.)

male, arousability immediately falls to zero following orgasm, while arousal decays slowly. These changes are illustrated in Figure 3.

At Time 0 sexual stimulation begins. Arousability is high, and arousal rapidly increases until orgasm occurs. At orgasm, arousability drops to zero for a short time and then slowly increases. After orgasm, arousal slowly declines if sexual stimulation is discontinued. At Time X after orgasm arousability would be lower than before orgasm but would be above zero, while arousal would still be relatively high. If the individual were restimulated at Time X, his rate of approach to a second orgasm would be slower than at the initiation of stimulation (Time 0), yet the second orgasm might occur with less total sexual stimulation than had the first orgasm, since the arousal level at the time of restimulation would be quite high. Again we see how important it is to distinguish arousal level from arousability. As noted earlier, the condition described here seems to hold for the male rat (Beach & Whalen, 1959a). Immediately following ejaculation, the male rat will not reinitiate mating. Five minutes after the ejaculation the arousal level is found to be quite high as measured by intromission latency, and the second ejaculation occurs after fewer intromissions than did the first ejaculation.

This analysis applies only to the male. There is no evidence that orgasm has similar feedback properties in the female. Masters (1960) has shown that orgasm does not necessarily alter arousability in women. Following an orgasm a woman may be maintained at a high "plateau" level of arousal and may be restimulated to a second orgasm in a short time. It is not known, however, whether the second orgasm is reached with more or less stimulation than the first. It is possible that in women arousal levels are maintained at such a high level after orgasm that changes in arousability are obscured.

We are proposing that sexual arousability in humans as well as in nonhuman animals does have a physiological basis. Arousability is influenced by hormones with their central and peripheral actions and by the feedback effects of sexual stimulation whether mediated by orgasm or by stimulus-adaptation processes. This position is in direct contrast to that of Hardy (1964) who feels that:

> It seems warranted to conclude that the overwhelming proportion of the variance in human sexual motivation and behavior is not explicable in terms of some biological need or tension, however conceived [p. 4].

While sexual arousability does have a physiological basis, it is also influenced by experience in the sense that sexual arousal comes to occur to an increasing number of stimuli with experience. As pointed out by Hardy (1964), the adolescent is less likely to be aroused by movie love scenes than is an adult who has experienced sexual gratification in the ways depicted. As the adolescent matures and experiences sexual gratification in a variety of situations, new stimuli come to take on sexually arousing properties.

When the individual later returns to these situations, more stimuli are effective in producing arousal than had been the case before conditioning, and the individual reaches orgasm more rapidly. The arousability of the individual would, therefore, be greater, the greater the number of stimuli which could be simultaneously effective. It should be remembered, of course, that experience may also attach negative values to stimuli such that some stimuli may inhibit rather than increase arousal. Arousability would be lowered in such situations.

SEXUAL GRATIFICATION

We are proposing that arousability is conditionable. If this is true, how is this conditioning effected? It is likely that stimulation of the erogenous zones is inherently reinforcing, and that neutral stimuli become conditioned arousers through their association with stimulation of the erogenous zones. The primary mode of reinforcement would be through genital stimulation as produced in the male by phallic contact or possibly as produced spontaneously during erection. That penile stimulation is inherently reinforcing has been suggested by work with lower animals (Beach & Jordan, 1956a; Bermant, 1961; Kagan, 1955; Whalen, 1961). That orgasm or ejaculation is not necessary for reinforcement to occur has also been demonstrated (Sheffield, Wulff, & Backer, 1951; Whalen, 1961).

In women, clitoral and breast stimulation are presumably reinforcing, but there is little evidence to confirm this point. The female rat will learn an instrumental response in order to mate with a male (Bermant, 1961) which suggests that some component of the mating pattern is inherently reinforcing.

It is clear that the stimulation provided by orgasm is more reinforcing than that provided by intromission or clitoral contact. Women who achieve orgasm by masturbation are more likely to continue the practice of masturbation than are women who do not achieve orgasm in this way (Kinsey, Pomeroy, Martin, & Gebhard, 1953). Male rats are more likely to perform an instrumental response to be with a female if they are allowed to ejaculate in the goal box than if only intromission is permitted (Kagan, 1955). These studies, however, do not show that orgasm has any unique reinforcing properties. It may be that the reinforcement characteristics of orgasm reflect only the summated reinforcement properties of the stimulation which is needed to produce orgasm.

SEXUAL IDENTIFICATION AND OBJECT CHOICE

The present paper has attempted to specify some of the biological and psychological influences upon human and animal sexual motivation. Students of human behavior, however, may find these statements irrelevant to what they consider the central problem of human sexual motivation, namely, sexual identification (or masculinity and femininity) and object choice, the selection of the persons or objects toward which sexual activities are directed. It is necessary, therefore, to relate these dimensions of sexuality to the motivational concepts which have been discussed.

Sexual Identification

Money, Hampson, and Hampson (1955a, 1955b, 1956), in an excellent series of studies, have shown that an individual's concept of his gender, male or female, is determined primarily by the way in which the individual was reared. Pseudohermaphrodites with identical gonads and anomalous genital development consider themselves and are considered by others either male or female depending upon whether they were raised as male or female. These findings seem to conflict with our knowledge of animal sexuality which indicates that the development of male and female sexual motivation is determined by genetic hormonal influences (Young, Goy, & Phoenix, 1964). This conflict arises because it is thought that human sexual identification allows as reasonable inferences about sexual motivation as does animal sexual behavior.

However, "sexual identification" and "gender role" are psycho- and sociosexual concepts denoting socially defined, sexually dimorphic (or more properly, sexually dipsychic) behaviors. Gender-specific behaviors vary widely from culture to culture and include activities such as working, cooking, hunting, and house cleaning, as well as the initiation of sexual activities (cf. Mead, 1961). The gender-role concept does not necessarily include the concepts of sexual motivation, arousal, or arousability. There is no evidence to indicate that sexual motivation is inherently dimorphic and characteristic of one gender or the other. Of course, the specific forms taken by sexual behavior are affected by the sexual identification of the individual. For example, the propensity for arousal by visual stimuli may be considered one component of the male gender role (Kinsey et al., 1953). However, one would not want to conclude from data of this nature that males and females differ in the type or form of erotic arousal in any meaningful way. We find it reasonable to assume, rather, that sexual arousal is similar in males and females, although the expression of arousal is sexually dimorphic and determined by an individual's experience within his culture.

Furthermore, while the sexual identification of an individual determines some facets of the expression of sexual arousal, not all components of sexual arousal are gender specific. It is not even clear, for example, that the copulatory responses of men and women are entirely distinctive. Certainly the stereotyped postural responses characteristic of nonhuman species are not found in humans. The little experimental work which has been done on human sexual behavior indicates that one aspect of the copulatory response pattern, orgasm, is essentially equivalent in males and females. Masters and Johnson (1960, 1963) have shown that the vaso-congestive phenomena and the pelvic-genital contraction phenomena which constitute orgasm are similar in males and females. Mechanical and fluid-expulsion differences are, of course, present, but these should not cloud the basic anatomic and physiologic similarities which are present. These similarities should be considered particularly striking since orgasmic processes have been considered quite different in males and females for so long.

It would seem, therefore, that the very useful concept of gender role or sexual identification should be considered orthogonal to other facets of sexuality such as sexual arousal and the index of maximal arousal, orgasm, and that only concepts of arousal and arousability need be considered motivational constructs.

Object Choice

Hardy (1964) notes that the frequency and mode of homosexual relationships vary widely from culture to culture and from age to age, and, therefore, cannot result from biological (hormonal) influences. Hampson and Hampson (1961) also dismiss the hormonal determination of object choice because they have repeatedly found individuals of similar hormonal status who direct their sexual activities to males or to females in a way which reflects their sex of rearing rather than their hormonal sex. These investigators conclude from these data that learning and experience, rather than hormones, determine the choice of sexual object. The implication from this conclusion is that learned object choice is part of the sexual motivation. This implication is unwarranted. There is certainly little evidence, even in the animal literature, to indicate that object choice determines sexual motivation, or that sexual motivation determines object choice. In animals as in man, object choice *may* be determined by experience. Dogs masturbated by an experimenter will come to direct their sexual behavior toward that experimenter (Beach, 1950), and male cats may be readily trained to mount other males or even toys (Michael, 1961). In these cases one would not say that motivation has been altered by the training experiences.

We feel that the object chosen is learned and represents one facet of the habit component of the expression of sexual behavior. This distinction is similar to the one drawn by Freud (1962).

APPROACHES TO SEXUAL MOTIVATION

At several points in this discussion, the present propositions have been placed in contrast to those of Hardy (1964). These contrasts have been drawn in order to highlight our differences in theoretical approach to the problem of sexual motivation.

Hardy has followed one historically valid approach to the problem of motivation, namely, one which considers the direction of behavior an integral part of the motivation. In such cases, motives are thought to act partially through a process of stimulus selection. The hungry organism orients toward food and not toward a mate. When applied to sexuality, such a theory leads to the notion that homosexuality and heterosexuality represent fundamentally different motivational states. The present theory, in contrast, may be characterized as "energetic" theory or "arousal" theory in which the expression of the motivation is considered independent of the motivational state. Homosexual and heterosexual behaviors, in the present theory, are considered different habit states which may be activated by identical motivational states.

These different theoretical approaches to sexual motivation lead to quite different relative emphases upon experience and the biological basis of sexuality. Hardy's response-directed theory finds little biological contribution to sexual motivation, since so much of the expression of human sexual behavior is determined by experience. The present energetic theory, on the other hand, emphasizes the biological basis of motivation and considers the expression of behavior a combined function of motivation and habit. Whether the response-directed theory or the energetic theory will prove more effective in eliciting fresh research and new theoretical formulations must remain an empirical question.

REFERENCES

BEACH, F. A. *Hormones and behavior.* New York: Hoeber, 1948.

BEACH, F. A. Sexual behavior in animals and man. *The Harvey Lectures,* 1950, Ser. 43, 259–279.

BEACH, F. A., & JORDAN, L. Effects of sexual reinforcement upon the performance of male rats in a straight runway. *Jour-*

nal of Comparative and Physiological Psychology, 1956, **49**, 105–110. (a)

BEACH, F. A., & JORDAN, L. Sexual exhaustion and recovery in the male rat. *Quarterly Journal of Experimental Psychology,* 1956, **8**, 121–133. (b)

BEACH, F. A., & WHALEN, R. E. Effects of ejaculation on sexual behavior in the male rat. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 249–254. (a)

BEACH, F. A., & WHALEN, R. E. Effects of intromission without ejaculation upon sexual behavior in male rats. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 476–481. (b)

BERMANT, G. Response latencies of female rats during sexual intercourse. *Science,* 1961, **133**, 1771–1773.

BERMANT, G. Effects of single and multiple enforced intercopulatory intervals on the sexual behavior of male rats. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 398–403.

BREMER, J. *Asexualization.* New York: Macmillan, 1959.

CARTER, A. C., COHEN, E. J., & SHORR, E. The use of androgen in women. In R. S. Harris & K. V. Thimann (Eds.), *Vitamins and hormones.* New York: Academic Press, 1947. Pp. 317–391.

FORD, C. S., & BEACH, F. A. *Patterns of sexual behavior.* New York: Harper, 1951.

FOSS, G. L. The influence of androgens on sexuality in women. *Lancet,* 1951, **1**, 667–669.

FREUD, S. Trans. by J. Strachey (Ed.), *Three essays on the theory of sexuality.* New York: Basic Books, 1962.

FULLER, W., & DREZNER, N. The results of surgical castration in women under forty. *American Journal of Obstetrics and Gynecology,* 1944, **47**, 122–124.

GROOME, J. R. A local use of testosterone. *Lancet,* 1939, **2**, 722.

HAMPSON, J. L., & HAMPSON, J. G. The ontogenesis of sexual behavior in man. In W. C. Young (Ed.), *Sex and internal secretions.* Baltimore: Williams & Wilkins, 1961. Pp. 1401–1432.

HARDY, K. R. An appetitional theory of sexual motivation. *Psychological Review,* 1964, **71**, 1–18.

HARLOW, H. F. The heterosexual affectional system in monkeys. *American Psychologist,* 1962, **17**, 1–9.

KAGAN, J. Differential reward value of incomplete and complete sexual behavior. *Journal of Comparative and Physiological Psychology,* 1955, **48**, 59–64.

KINSEY, A. C., POMEROY, W. B., MARTIN, C. E., & GEBHARD, P. H. *Sexual behavior in the human female.* Philadelphia: Saunders, 1953.

KUEHN, R. E., & BEACH, F. A. Quantitative measurement of sexual receptivity in female rats. *Behaviour,* 1963, **21**, 282–299.

KUPPERMAN, H. S., & STUDDIFORD, W. E. Endocrine therapy in gynecologic disorders. *Postgraduate Medicine,* 1953, **14**, 410–424.

LARSSON, K. *Conditioning and sexual behavior in the male albino rat.* Stockholm: Almquist & Wiksell, 1956.

MASTERS, W. H. The sexual response cycle of the human female: I. Gross anatomic considerations. *Western Journal of Surgery, Obstetrics, and Gynecology,* 1960, **68**, 57–72.

MASTERS, W. H., & JOHNSON, V. E. The sexual response of the human male: I. Gross anatomic considerations. *Western Journal of Surgery, Obstetrics, and Gynecology,* 1963, **71**, 85–95.

MEAD, MARGARET. Cultural determinants of sexual behavior. In W. C. Young (Ed.), *Sex and internal secretions.* Baltimore: Williams & Wilkins, 1961. Pp. 1433–1479.

MICHAEL, R. P. "Hypersexuality" in male cats without brain damage. *Science,* 1961, **134**, 553–554.

MONEY, J. Components of eroticism in man: I. The hormones in relation to sexual morphology and sexual desire. *Journal of Nervous and Mental Disorders,* 1961, **132**, 239–248.

MONEY, J., HAMPSON, J. G., & HAMPSON, J. L. An examination of some basic sexual concepts: The evidence of human hermaphroditism. *Bulletin of the Johns Hopkins Hospital,* 1955, **97**, 301–319. (a)

MONEY, J., HAMPSON, J. G., & HAMPSON, J. L. Hermaphroditism: Recommendations concerning assignment of sex, change of sex and psychologic management. *Bulletin of the Johns Hopkins Hospital,* 1955, **97**, 284–300. (b)

MONEY, J., HAMPSON, J. G., & HAMPSON, J. L. Sexual incongruities and psychopathology: The evidence of human hermaphroditism. *Bulletin of the Johns Hopkins Hospital,* 1956, **98**, 43–57.

RABEDEAU, R. G., & WHALEN, R. E. Effects of copulatory experience on mating behavior in the male rat. *Journal of Comparative and Physiological Psychology,* 1959, **52**, 482–484.

ROSENBLATT, J. S., & ARONSON, L. R. The decline of sexual behavior in male cats after castration with special reference to the role of prior sexual experience. *Behaviour,* 1958, **12**, 285–338.

SALMON, U. J., & GEIST, S. H. Effect of androgens upon libido in women. *Journal of Clinical Endocrinology,* 1943, **3**, 235–238.

SHEFFIELD, F. D., WULFF, J. J., & BACKER, R. Reward value of copulation without sexual drive reduction. *Journal of Comparative and Physiological Psychology,* 1951, **44**, 3–8.

WAXENBERG, S. E., DRELLICH, M. D., & SUTHERLAND, A. M. The role of hormones in human behavior: I. Changes in female sexuality after adrenalectomy. *Journal of Clinical Endocrinology and Metabolism,* 1959, **19**, 193–202.

WHALEN, R. E. Effects of mounting without intromission and intromission without ejaculation on sexual behavior and maze learning. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 409–415.

WHALEN, R. E. The initiation of mating in naive female cats. *Animal Behavior,* 1963, **11**, 461–463. (a)

WHALEN, R. E. Sexual behavior of cats. *Behaviour,* 1963, **20**, 321–342. (b)

WINOKER, G. Aspects of sexual behavior: A classification. In G. Winoker (Ed.), *Determinants of human sexual behavior.* Springfield, Ill.: Charles C Thomas, 1963. Pp. vii–viii.

WITTENBORN, J. R. Inferring the strength of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957.* Lincoln: Univer. Nebraska Press, 1957. Pp. 191–259.

YOUNG, W. C., GOY, R. W., & PHOENIX, C. H. Hormones and sexual behavior. *Science,* 1964, **143**, 212–218.

(Received November 9, 1964)

Psychological Review
1966, Vol. 73, No. 2, 164–178

STIMULUS VARIABILITY AND COGNITIVE CHANGE[1]

HARRY MUNSINGER
University of Illinois

AND

WILLIAM KESSEN
Yale University

4 postulates of a tentative theory of cognitive change are presented. Implications of the postulates for response to stimulus variability as a function of age were tested in 4 studies ($N = 863$). Measures of response used were estimation accuracy, the ability to learn class names, the ability to classify, and changes in expressed preference after differential experience with stimulus variability. In each study children at 2 grade levels and adults responded to random shapes varying from 5–40 independent turns. Results supported the following conclusions: (a) Ss are sensitive to variability; (b) there is a limit on processing ability; (c) there are systematic effects of experience with variability on expressed preference; (d) young children tend to select from the presented variability. The postulation that experience with stimulus variability just beyond the limits of processing ability would result in maximal cognitive change received partial support.

Since the publication of Hebb's *The Organization of Behavior* in 1949, psychologists have become increasingly concerned with two closely related problems—the definition of cognitive structure and the specification of a mechanism for cognitive change.

There is great diversity in the definition of cognitive *structure*. Hebb's emphasis has been on the elaboration of perceptual processes as a general model for cognition. Bartlett (1932) discussed schemata as the structures underlying memory and its distortions, while Piaget (1950) used the same word to comprehend a very different range of phenomena—the development of intelligent behavior in children. Two more recent statements have extended the notion that human thought can be conceived as a system of rules. Newell, Shaw, and Simon (1958) discuss the application of computer technology and logic to an understanding of human thinking, and Miller, Galanter, and Pribram (1960) propose the mechanism Test-Operate-Test-Exit as the basic unit of cognitive structure. The theme most clearly common to these and other attempts to encompass human thought in structural terms is their rejection of the topographically defined response as a model for cognition. In language that varies from metaphor to mathematics, cognitive theorists propose that man's thought is a system of rules for processing environmental change, for building representations of things, and for ordering these representations to suit the demands of the person and his world.

Variation of opinion about the nature of cognitive *change* also contains a common rejection; almost without exception, cognitive theorists have denounced the use of physiological deficit and its derivatives in explaining the development of human thought. A number of mechanisms have been proposed as substitutes for this traditional view of motivation. Berlyne (1960) has suggested curiosity drive, and Butler (1953) discusses stimulus hunger,

---

[1] The research reported here was supported in part by a grant from the Carnegie Corporation of New York.

but most recent proposals have contained the notion of discrepancy between sensory input and cognitive structure or discrepancy between two elements of cognitive structure as demanding cognitive change. The removal or reduction of a discrepancy is assumed to be responsible for cognitive change (see Hunt, 1963, for a review of several theoretical positions on this issue).

Even before questions about cognitive structure and change can be addressed, it is necessary to make some statement about the nature of the environment from which cognitive structures must develop. Some psychologists have continued to use the traditional division of the environment into common objects and, rarely, physical quantities under the general term *stimulus* with occasional reference to characteristics of receptor end organs. A somewhat more general procedure for description of the environment has been suggested in the work of Attneave (1957) and Garner (1962), among others. For these psychologists, the environment is seen as a field of variability characterized by relative independence or redundancy of events. They have used modifications of communication theory (Shannon & Weaver, 1949) to develop metrics for the specification of environmental variability.

We have recently reported (Munsinger & Kessen, 1964; Munsinger, Kessen, & Kessen, 1964) a series of studies on the structuring of environmental variability. Specifically, we asked subjects to state their preferences for (*a*) random shapes that varied widely in number of independent elements, and (*b*) sequences of letters and words that varied from random to fully redundant. Those studies and the ones to be reported here grew out of several general assumptions. The purpose of the present report is to state these governing assumptions briefly, to present four new studies derived from them, and, in the light of the results of these studies, to indicate required modifications and extensions of the assumptions. Our initial assumptions were four.

1. *Variability.* Human beings are sensitive to stimulus variability; this variability may be usefully specified in the metric of information theory (Attneave, 1959).

2. *Limit.* There is an upper limit on the capacity of human beings to process independent environmental events (Miller, 1956).

3. *Structuring.* Human beings are able to break through this limitation on their capacity to process independent events through the development of rules (structure) which reduce the effective stimulus variability (Mandler, 1962).

These first three assumptions can be interrelated in another way. People respond to cognitive uncertainty. In the limiting case, the unrealistic one in which there is no organizing cognitive structure, cognitive uncertainty is a simple function of stimulus variability. However, even for unsophisticated human beings—the newborn infant, at the extreme—rules are available which permit some coding or processing of stimulus variability. Thus we can think of cognitive uncertainty as a joint function of cognitive structure and stimulus variability.

It should be remembered that, in the present studies, we have assumed that human beings have a preference for a certain level of cognitive uncertainty (Munsinger & Kessen, 1964). Cognitive uncertainty reflects the person's ability to impose meaning (cognitive structure) on the array of information available in his environment (stimulus variability). If stimulus variability changes with no change in cognitive structure, then preference should shift. Similarly, if the subject imposes structure on a fixed field of stimulus variability, then his preference will change also. Certainly preference is *not* the only index of cognitive uncertainty; for this reason we have reported relevant evidence on estimation accuracy and categorization. The measure of preference, however, has the advantage of expressing, in part, the motivational aspects of cognitive uncertainty. If the development of structure results in the subject's preference for increasing amounts of stimulus variability, we have a general model for intellectual development. As the person learns to structure variability, he turns toward higher levels of variability.

It is not altogether clear how rules are imposed on random shapes to reduce stimulus uncertainty. Subjects may categorize the figures in terms of previous experience ("They look like something.") or, in the case of the art-student subjects, in terms of an imposed dimensionability (jaggedness or weight), or in terms of specific memorization. However the structuring takes place, it seems clear that it is accompanied by changes in accuracy of estimation, accuracy of categorization, or preference.[2]

4. *Cognitive Change.* A person will be most likely to change his cognitive structure if he is presented with a level of cognitive uncertainty just beyond his present processing ability. If he is presented cognitive uncertainty far beyond his ability to process or no cognitive uncertainty at all, little or no change will take place (Piaget, 1952).

In our first studies derived from these assumptions we decided to examine human preference for variability. In part, this decision grew from our concern with problems of motivation. What aspects of the environment would claim the attention of a person? What parts of the environment would be attractive to him? As Piaget (1952) says ". . . a subject looks neither at what is too familiar, because he is in a way surfeited with it, nor at that which is too new, because this does not correspond to anything in his schemata . . . [p. 68]." Thus, in earlier studies of the structuring of variability, we examined the specific hypothesis that subjects would indicate a preference for random shapes in the middle range of stimulus variability, and that this level of preference could be changed with the development of rules for coding stimulus variability. These results will be reviewed in the following pages as we examine new evidence on three aspects of the formulation presented here—the limitation on processing ability, the development of cognitive structure, and the role of specific experience in forming cognitive structure.

PROCESSING LIMITS

Miller (1956) has reviewed much of the earlier literature on the existence of an upper limit on the ability of human beings to receive, process, and transmit environmental variability. Garner (1962) has summarized more recent literature on this issue. In studies of preference and variability we found an inflection in the preference-variability function with stimuli of about 10 independent turns. Although such findings fit well with the earlier literature on processing limits and with our assumptions, there was a clear need for an independent test of the assumption of the processing limit. Study I was designed to provide such a test.

---

[2] Another interesting example of change in behavior with the assumed development of structure is provided by a study of estimation accuracy in which two groups of subjects ($N = 120$) were asked to look at and to try to remember 5-turn figures (Group 1) or 40-turn figures (Group 2). After 140 presentations at 7 seconds duration, all subjects were asked to estimate four levels of stimulus variability at 200 milliseconds exposure duration. A statistically stable ($F_{8/340} = 4.01$; $p < .01$) interaction reflected the fact that Group 2 was more accurate in estimation of 40-turn figures.

Study I

Subjects. Fifty-four undergraduate students in psychology courses at the University of Illinois and 108 students in the second, third, fifth, and sixth grades of School District Number 68, Skokie, Illinois, served as judges in this study. Data from the second and third graders were combined, as were data from the fifth and sixth graders.[3]

Stimulus Materials. Seven instances of eight levels of stimulus variability represented by random figures of 5, 6, 8, 10, 13, 20, 31, and 40 independent turns (Attneave & Arnoult, 1956; Munsinger & Kessen, 1964) were presented to the judges in two randomized blocks. Each stimulus was shown on a projection screen for ½ second; the interval between presentations was approximately 10 seconds.

Procedure. Judges were shown examples of the figures and told the number of categories to be judged and the number of turns defining each category. They were asked to guess the number of turns in the figures projected. The procedure was essentially that used by Kaufman, Lord, Reese, and Volkmann (1949).

Results

Figure 1 shows the relation between number of independent turns defining each category of random shapes and judges' accuracy of estimation. The number of correct estimations decreased linearly between figures of 5 turns and figures of 10 turns. For figures beyond 10 turns, no further decrease in accuracy occurred ($F_{7/1,113}$ for linear trend = 639.0; $F_{7/1,113}$ for quadratic component = 68.6). More recent data from our laboratories indicate that, for relatively brief exposures, there is no interaction between duration of exposure and the limits of estimation accuracy. There is an apparent age difference in accuracy ($F_{2/159} = 16.2$; $p < .001$) and an interaction ($F_{14/113} = 8.3$; $p < .001$) between age and stimulus variability. The slopes of the functions beyond 10 turns are parallel from one age group to another.

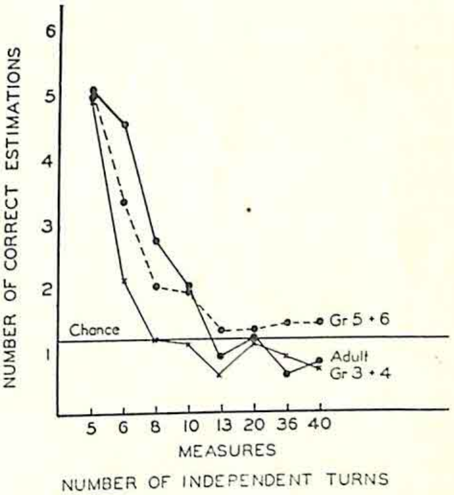

FIG. 1. Mean number of correct estimations of number of turns of eight levels of random shapes for three age groups.

These findings lend strong independent support to the postulate that there is an upper limit on the ability of human beings to process variability of stimulation. For the random shapes this limit is in the range of 8–10 independent turns. This finding is in good agreement with previous results which indicate an inflection in the preference-variability function at about 10 independent turns.

The relation between estimation accuracy and stimulus variability in the range between 5 and 10 independent turns that was found in the present study is reflected in the preference data of a number of earlier studies. For children from first to eighth grade and for adult subjects who have had varying experience with random shapes, the preference functions for figures between 5 and 10 independent turns are essentially parallel. This is not at all the case if we consider the stated preference of subjects for figures with more than 10 independent turns. The preference function for figures between 10 and 40 turns does not resemble closely the corresponding portion of the estimation-accuracy function. Moreover, the preference function for figures of high variability shows systematic and significant variation among groups of subjects who differ in age and in experience with random shapes (Munsinger & Kessen, 1964). This wide variation in response to figures above 10 turns suggests the operation of a procedure for processing high stimulus variability which is very different from the simple perceptual one which may be reflected by the results of the estimation study.

---

[3] The authors are grateful to Wesley Gibbs and the staff of Skokie School District Number 68 for their support and assistance.

THE DEVELOPMENT OF STRUCTURE

We have examined three sources of variation in the handling of figures of high variability—long-term experience with complex forms, short-term experience with the random shapes, and age (Munsinger & Kessen, 1964; Munsinger et al., 1964).

Long-Term Experience with Complex Forms

Art students were asked to state their preferences for random shapes varying from 5–40 independent turns in the expectation that their experience with geometrical figures would have led to the development of rules for structuring high variability. The results confirmed the expectation; the preference-variability function for the art students was generally monotonic and positive. It is interesting also that the art students found figures of few turns "dull, plain, and uninteresting."

Short-Term Experience with Random Shapes

A group of college undergraduates was asked to state their preferences over and over again for the same figures differing in stimulus variability under the hypothesis that increasing experience with random shapes would lead to the development of procedures for structuring. Preference for figures of high variability was expected to shift upward following 2 hours of paired-comparison preference judgments. Again, the results supported the conclusion that preference for high variability may be changed through development of coding rules. The subjects showed higher preference for figures of many turns toward the end of the session than they had at the beginning.

Age

The changing preferences of children for figures of high variability were explored by means of the same procedures for determining preference used with adults. Children in all grades of an elementary school made preference judgments of the 5- to 40-turn random shapes. The expectation of an age-invariant area of inflection of the preferance-variability function near figures of 10 independent turns was confirmed. There is surprising similarity in the shape of the preference-variability function between 5 and 10 independent turns when the responses of human beings from ages 6–22 are compared.

As has already been suggested, no such developmental invariance exists in expressed preference for figures of higher variability (i.e., figures of more than 10 turns). We had expected that children would have less ability to structure figures of high variability and that they would therefore show lower preference for them. However, the children prefer figures of high variability far more than adults do. The decreasing preference for figures of high variability with increasing age is regular; second graders state lower preferences for figures of high variability than do first graders, third graders state lower preferences for figures of high variability than do second graders, and so on. These unexpected findings lead to the studies of developmental variation in ability to categorize reported here.

Sense may be made of the unpredicted preference of young children for figures of high variability if it is assumed that young children and adults process high variability in the following different ways. Young children may sample the high-variability figures, attending only to that part of the figure which they can structure; figures of many turns provide more opportunity for this process of selection than do figures of few turns. Thus, young children could handle and would prefer figures of high variability. Adults, on the other hand, are more likely to attend to the entire figure whatever its level of variability. This strategy places a demand on adults to structure figures of many turns; their inability to do so leads unsophisticated adults to express relative dislike for figures of high variability.

A clear and testable proposition derives from this assumption. If young children attend to only part of the figures of many turns, then they should exhibit differential difficulty in categorization—that is, the formation and use of classes based on high stimulus variability—when their ability to categorize is compared with the ability of older children and adults. Moreover, younger children should show less improvement over trials in their ability to categorize than older children and adults because strategies of categorization based on selection of part of a random figure are inappropriate for the task of classification.

Two studies were designed to test this prediction about the relation of high stimulus variability and age to the ability to categorize. Study II required subjects to associate nonsense names to classes of random shapes of four levels of variability; Study III required subjects to sort random shapes on the basis of their variability.

Study II

Subjects. Twenty-two children from each of four different grade levels (second, third, fifth, and sixth) in Champaign, Illinois, schools and 22 adults from introductory psychology classes at the University of Illinois participated in Study II.[4]

Specific Method. Six different examples of each of four levels of stimulus variability were used. One level contained only random figures of 5 independent turns, another level only 10-turn figures, a third only 20-turn figures, and the fourth level contained only 40-turn figures. Subjects were told: "Try to learn which name goes with which figure. There are things about the figures which will help you to know which name goes with which figure." Subjects were encouraged to guess; they wrote their responses on prepared answer sheets.

Each figure and its "name" were projected for 10 seconds to a front wall of the room in which subjects were seated. The 24 different random shapes were presented with order randomized within balanced quarters on each of 6 runs. On the first, third, and fifth runs, the figures were followed by their appropriate names (nonsense syllables of matched medium-association value). On the second, fourth, and sixth runs, the figures appeared in a different order, and the names did not appear. Subjects were asked to write the appropriate name of each shape as it was presented for 10 seconds. The scores used in a mixed

---

[4] The authors are grateful to Florence Aye and her staff at the Dr. Howard Elementary School, Champaign, Illinois, for their support and assistance.

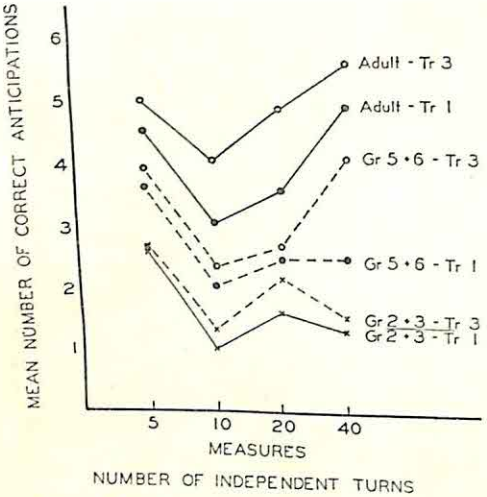

FIG. 2. Mean number of correct responses in classification of random shapes as related to age, stimulus variability, and trials.

analysis were the number of correct names produced by each subject for each level of variability. For all groups, 5-turn figures were named BIP, 10-turn figures were named QOR, 20-turn figures were named ZAL, and 40-turn figures were named TUJ. A response was scored as correct if the subject gave two of the three letters in the syllable in proper order. Two additional groups of subjects (22 fourth graders and 22 third graders) were run with different relations of nonsense syllables and figures. The results of this control series were the same as those described below.

Results

There was a consistent monotonic relation between age and accuracy of class naming ($F_{4/105} = 30.4$; $p < .001$). There were also statistically significant effects of trials—Trial 1 versus Trial 3—($F_{1/109} = 27.1$; $p < .001$) and stimulus variability ($F_{3/315} = 24.5$; $p < .001$). Of greatest interest, the predicted interaction among age level, stimulus variability, and trials shown in Figure 2 was statistically stable ($F_{12/315} = 2.31$; $p < .01$).

The results shown in Figure 2 demonstrate that younger children have differentially greater difficulty learning to associate names to classes of figures of many turns and, further, that younger children profit relatively less from experience. This effect is shown best by the differential achievement of correct names for the class of 40-turn figures. The second- and third-grade children have not learned the correct name after Trial 1, and the second- and third-grade children still have not learned the name of the class of many-turn figures after Trial 3. The fifth- and sixth-grade subjects respond with many more correct names to the 40-turn figures on Trial 3 as compared to their performance on Trial 1, while adults show relatively high attainment of correct solutions on both Trial 1 and Trial 3.

Study III

A second study of categorization was designed to test, in a somewhat different way, the assumption that younger subjects do not attend to all the presented variability in the environment but instead adopt a strategy of sampling. Groups of subjects at three age levels were asked to sort shapes of four different levels of variability into four piles of equal size. It was expected that younger subjects would show differentially higher error rates for shapes of high variability.[5]

Stimulus Materials. Photographs of 32 different random shapes were printed on 5 × 7 inch glossy paper. Two prints each of eight 5-turn figures, eight 10-turn figures, eight 20-turn figures, and eight 40-turn figures were used.

Subjects. Thirty-seven students in introductory psychology classes at the University of Illinois, and 37 third-grade and 37 sixth-grade pupils served as subjects.[6]

---

[5] The data for this study were collected by Marvin Daehler during his tenure as an NSF Undergraduate Research Student.

[6] The authors are grateful to Floyd Daub and his staff at Milledgeville Elementary School for their support and assistance.

Procedure. Subjects were given the stack of 64 photographs in random order and the following instructions: "Here is a set of pictures. Sort them into four groups which you think seem to go together. Make each of the piles approximately equal in size. Don't take too long looking at the pictures, just look at one and put it into one of the piles. You may look at several pictures before beginning to sort them into piles. Any questions? OK, you may begin."

Results

Percentage of error was calculated for all subjects for the four levels of variability of shapes. To correct for response bias the error score used for each category was $N-/N_T$; that is, number of shapes appropriate to that category which were placed outside it ($N-$) divided by the total number of shapes, whether appropriate or not, placed in the category (N_T). The simpler measure of number of correct placements in each category produced essentially the same results. Figure 3 presents the results of Study III. The ordinate represents a reciprocal form of the error measure in order to make the results shown comparable to the results obtained in Study II. The younger the subject, the more difficulty he has in correctly categorizing figures of many turns. This interaction between age and stimulus variability is significant ($F_{6/324} = 3.37$; $p < .01$). The effect of age is significant ($F_{2/108} = 10.3$; $p < .001$), as is the effect of stimulus variability ($F_{3/324} = 43.7$; $p < .001$).

An interesting secondary finding resulted from an analysis of the direction of miscategorization. When only shapes erroneously categorized were considered, shapes rated *high* in meaningfulness were placed in the category *below* the correct one, and shapes rated *low* in meaningfulness were placed in the category *above* the correct one ($\chi^2 = 4.1$; $p < .05$). This finding conforms with our earlier observation that meaningfulness reduces the apparent variability of random shapes (Munsinger & Kessen, 1964).

The results of these two studies of categorization are congruent with the notion that young children do not attend to all the variability presented in figures of many turns. In both studies there was a reliable interaction between age and level of stimulus variability with almost all of the interaction variance contributed by age variation in the categorization of 20- and 40-turn figures. Thus, regardless of the overall shape of the function, there seems to be a clear difference across age in the ability to classify figures of high variability. However, neither our argument nor the data in support of it are conclusive; the preference of children for high-variability figures remains somewhat puzzling.

The studies of categorization, while they help us to understand developmental changes in the structuring of variability, introduce additional prob-

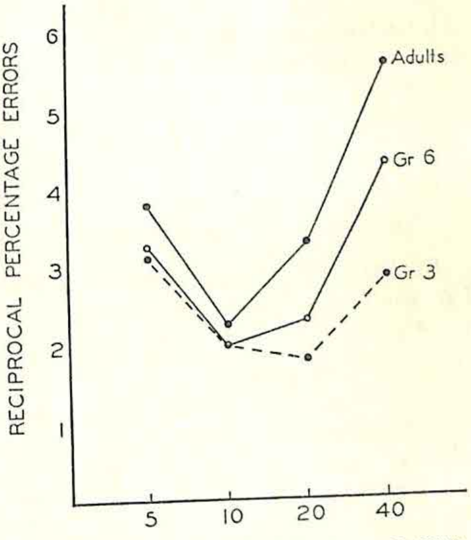

FIG. 3. Reciprocal percentage errors made by subjects of different ages in categorizing figures of different numbers of independent turns.

lems for the theoretical position we have taken. How is it that adults are able to categorize 40-turn figures so well, in the light of our earlier conclusions that 40-turn figures should be difficult for adults to structure? This formation of the question leads to a clarification of what is meant by the structuring of variability. The ability of an adult to differentiate a particular figure of high variability from others or even his ability to classify figures of high variability does not entail his understanding of the figures, that is, his ability to make sense of them in terms of his present cognitive structure. To take a different example, one can recognize the differences between classical mechanics and relativity without being able to structure relativity. It is for this reason that the implications of the categorization studies are equivocal. The fact that adults are able to categorize figures of high variability does not imply that they can structure them. On the other hand, the inability of young children to categorize figures of high variability lends support to, but does not demonstrate, the conclusion that young children respond to figures of high variability by sampling only part of the presented variability.

EXPERIENCE AND STRUCTURE

Perhaps the most interesting problem for a theory of cognitive structure is the specification of a mechanism of cognitive change. As we have seen, a number of commentators have pointed toward the importance of stimulation which is just beyond the person's present ability to structure. This position concerning cognitive change is worked out in greatest detail in the observations Piaget made on his children during their first 3 years of life. Having specified a domain for cognition — stimulus variability — and having found systematic variation of human response in this domain, it is possible for us to state hypotheses about the conditions for cognitive change. Let us say that a person expresses a preference for figures of intermediate variability. This expression of preference indicates that intermediate variability is just beyond his present ability to structure. If we want to shift preference by subjecting him to a range of experience that will be optimal for cognitive change, it would appear that (*a*) stimulation within his present level of preference would have little or no effect on his preference structure; (*b*) stimulation well above his present preference level would also have little effect on his preference structure because such stimulation is far beyond his present capacity to structure; and (*c*) stimulation in the range just beyond his present maximal preference should be optimally effective in changing cognitive structure and preference. Because of the apparent selective response of children to figures of high variability, it is difficult to predict the effect in children of experience with figures of high variability.

Study IV

In Study IV, second and third graders, fifth and sixth graders, and adults were given restricted experience with four levels of stimulus variability. Change in expressed preference for random shapes was used as a measure of the effect of experience with variability. As a further test of the effect of such experience, subjects were tested either immediately after experience with one of the levels of variability, or 24 hours later. This study was designed, therefore, to assess the role of restricted experience with variability in producing cognitive change, and the variation of these effects over age and time.

Subjects. A total of 480 subjects served in this experiment. Three hundred and twenty elementary school children from Champaign, Illinois, schools [7] and 160 college students from the University of Illinois undergraduate psychology subject pool served as subjects. One hundred and sixty second- and third-grade pupils were in the "young" age group, and 160 fifth- and sixth-grade pupils were in the "intermediate" age group. Both groups had equal numbers of boys and girls and equal numbers of subjects from each component grade.

Stimulus Materials. Five sets of eight random shapes were constructed using a procedure developed by Attneave and Arnoult (1956) and described by Munsinger and Kessen (1964).

One set of eight random shapes contained figures of 5, 6, 8, 10, 13, 20, 31, and 40 independent turns. This set of shapes was called the Test Set. Each of the other four sets contained eight different shapes of the same number of turns. Thus, the 5-turn set was made up of eight different figures of 5 independent turns, the 10-turn set of eight 10-turn figures, the 20-turn set of eight 20-turn figures, and the 40-turn set of eight 40-turn figures.

The stimuli were projected in pairs to the front wall of the room in which the subjects were seated. Subjects were instructed to state their preferences for each pair of stimuli.

Procedure. Eight groups of 20 subjects were used at each of the three age levels. Subjects in each age group were first asked to state their preferences for the Test Set of random shapes presented in an optimal order (Ross, 1934) according to a paired-comparison procedure. This allowed an estimation of a base level of preference for all the groups. Immediately following the presentation of Test Set for judgment, two randomly selected groups of subjects at each age were presented four successive complete paired-comparison preference runs (112 judgments) of 5-turn random shapes. Two other randomly selected groups, immediately after completing judgments of Test Set, were presented four successive paired-comparison preference runs with the 10-turn random shapes. Two of the remaining groups, immediately after completing judgments of Test Set, were presented four successive paired-comparison runs with 20-turn shapes. The remaining two groups, immediately after completing judgments of Test Set, were presented four successive paired-comparison preference runs with the 40-turn random shapes.

Thus each group within each age level judged a Test Set of varied random shapes (28 pairs of stimuli) and then made 112 judgments of pairs of 5-turn, 10-turn, 20-turn, or 40-turn random shapes. The three age groups were run at different times. For each age level, following the repeated exposure to one of the four different levels of variability experience, one of the two groups from each experience level was asked to leave the room and to return again the next day. The four remaining groups of 20 subjects at each age level immediately made paired-comparison judgments of the same Test Set seen previously. The other four groups of 20 subjects returned 24 hours later and made paired-comparison judgments of the same Test Set of random shapes.

In summary a total of 24 groups was used. Four groups at each of three age levels were tested for change in preference for variability immediately after experience with either 5-, 10-, 20-, or 40-turn random shapes. Four other groups at each of three age levels were given comparable experiences with levels of variability, but were tested 24 hours later.

Preference data were analyzed by conversion of pairwise preference to normalized deviates and then by taking least-squares solutions for the complete matrix of the resulting deviates as was done in earlier studies (Munsinger & Kessen, 1964).

Preference scores for the Test Set replicated the results of earlier studies in showing high preference for an intermediate level of variability (10 turns). Because our interest was in change of preference, difference scores were used in the analysis. A mixed-design analysis of variance was performed on the difference scores between judgments made during the first presentation of Test Set and judgments made during the second presentation of Test Set. The response measure used to compute the difference score was the number of "votes for" each random shape in Test Set by each subject. To ensure unconstrained row variance, votes for only the 5-, 10-, 20-, and 40-turn shapes of the Test Set were included in the analysis.

---

[7] The authors are grateful to Anne Mallory and her staff at Bottenfield Elementary School, Champaign, Illinois, for their support and assistance.

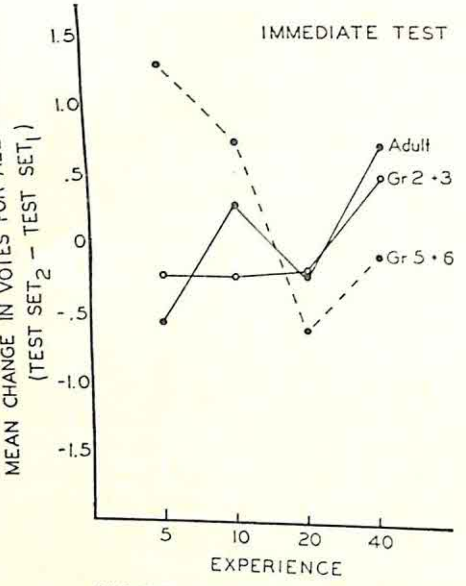

Fig. 4. The change in preference for random shapes of subjects at three age levels immediately after experience with levels of variability.

Results

It should be noted that figures of varying numbers of turns are used in two distinct ways—as experimental manipulation (*Experience*) and as response measure (*Measure*). Figures 4 and 5 show the interaction among age, experience with variability, and time of testing summed over the four measures. A between-groups analysis of variance shows this multiple interaction to be highly significant ($F_{6/456} = 10.2$; $p < .001$). As can be seen, neither adults nor *young* children markedly change their preference for variability immediately after the experimental manipulation of experience (Figure 4); the children of intermediate age change substantially, particularly in their response to experience with 5- and 10-turn figures. Twenty-four hours after experience

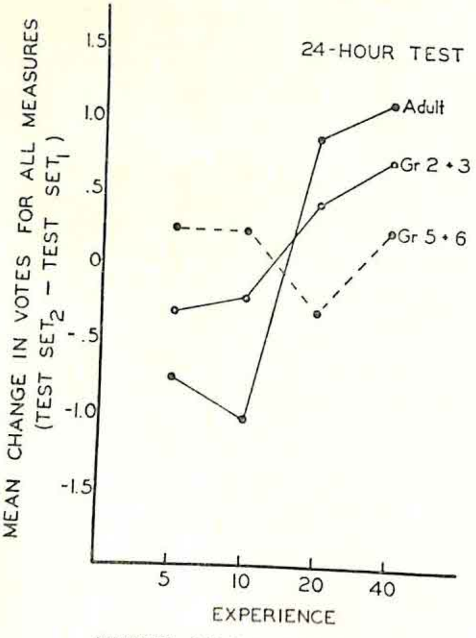

Fig. 5. The change in preference for random shapes of subjects of three age levels 24 hours after experience with levels of variability.

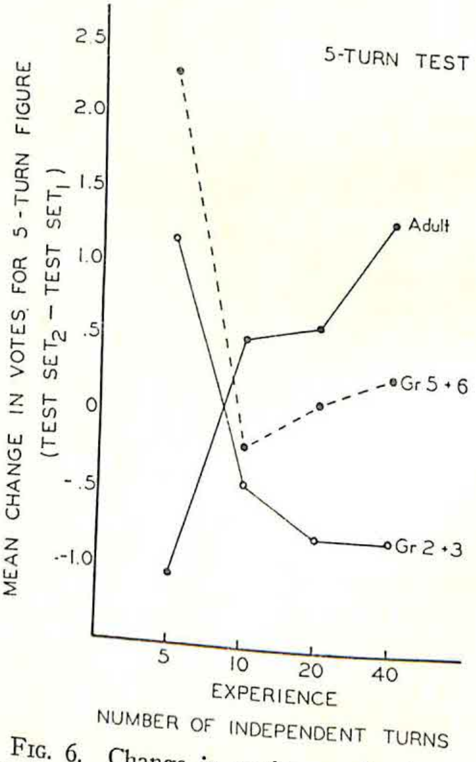

Fig. 6. Change in preference for figures of 5 turns of subjects at three age levels after experience with levels of variability.

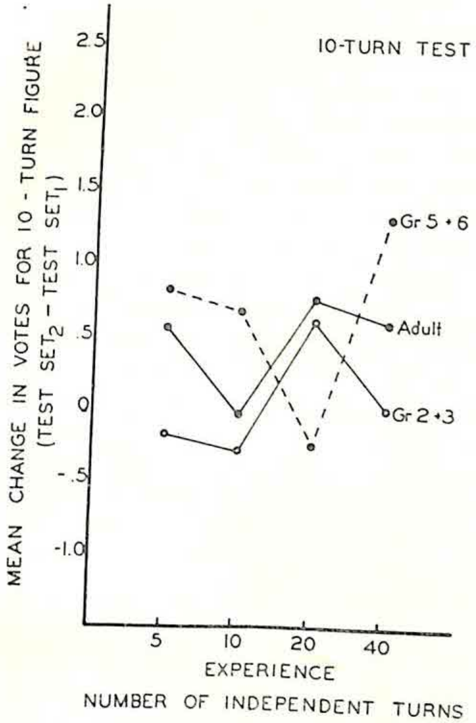

FIG. 7. Change in preference for figures of 10 turns of subjects at three age levels after experience with levels of variability.

with variability (Figure 5), children of intermediate age show relatively little effect of the manipulation, young children shift somewhat, and adults change their expressed preference dramatically.

The effects of experience with variability can be further understood by a consideration of the changes in expressed preference for the *different measures* at the three age levels. Changes in preference for the four measures are plotted separately in Figures 6, 7, 8, and 9. Statistical evaluation of the *measures* requires within-subjects comparison. The interaction of Measures × Age × Experience × Time reflected relatively well the time difference shown in Figures 4 and 5, but did not reach statistical significance. The interaction shown in Figures 6, 7, 8, and 9—Age × Experience × Measures—was significant ($F_{18/1.368} = 2.25$; $p < .01$).

Subjects of all ages showed relatively little differential change in their expressed preference for figures of 10 and 20 turns (Figures 7 and 8). In general, 10-turn figures were liked somewhat better in the second presentation of the Test Set, and 20-turn figures somewhat less. The interesting differences, both across experience and across age, were in response to the 5-turn shapes (low variability), shown in Figure 6, and to the 40-turn shapes (high variability), shown in Figure 9. In Figure 6 (response to 5-turn figures as measure) the young and intermediate-age children show much the same pattern of shifts in preference. Experience with 5-turn figures increased their expressed preference for 5-turn figures. The other experience conditions produced little effect among

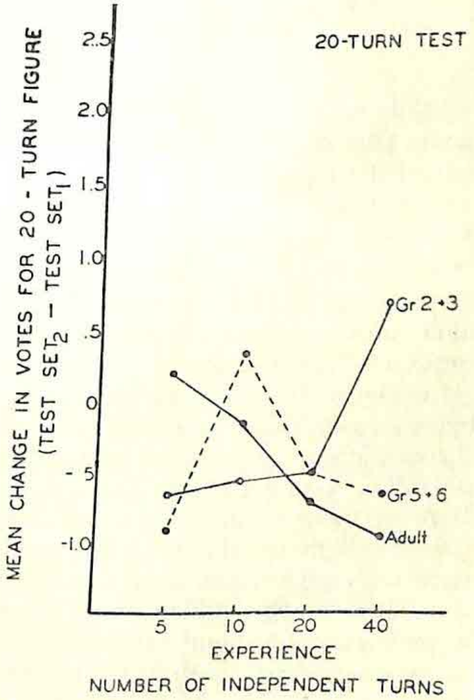

FIG. 8. Change in preference for figures of 20 turns of subjects at three age levels after experience with levels of variability.

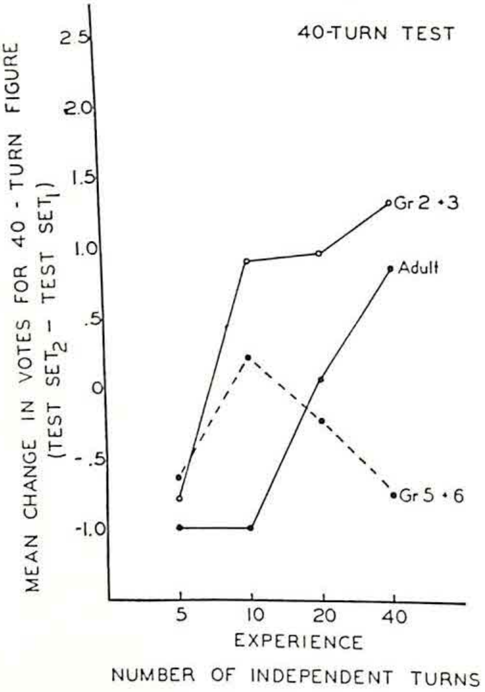

FIG. 9. Change in preference for figures of 40 turns of subjects at three age levels after experience with levels of variability.

the children. Adults disliked 5-turn figures after experience with them, but somewhat strangely turned to the low-variability figures after 40-turn experience. This shift represents largely an *immediate* reaction of the adults; 24 hours after 40-turn experience the adults do not show an increased preference for 5-turn figures.

In Figure 9 (response to 40-turn figures as measure) the similarity of changes in expressed preference is between the young children and adults. Both of these groups shift in preference for high variability after experience with high variability. In addition, the young children also show a preference for high variability following experience with both 10- and 20-turn figures. In contrast, the intermediate-age children shift their expressed preference away from high variability under all conditions except 10-turn experience.

The results of Study IV show systematic effects of differential experience with variability on preference judgments, but the changes in behavior cannot be interpreted under the simple postulate about cognitive change proposed earlier. A closer examination of Figures 6 and 9 will indicate where our theoretical presuppositions make sense and where they are in need of adjustment. The behavior of adults in response to 40-turn test stimuli is congruent with expectation; experience with high variability (20- and 40-turn figures) results in an increased ability to process variability and a corresponding shift in preference. For young children, a parallel performance must be given a different interpretation. Taking these responses of second and third graders together with the results of Studies II and III, one may conclude that experience with 10-, 20-, and 40-turn figures gives young children an opportunity to apply their sampling-of-variability tactic. They find a variety of processable samples in figures of high variability, and they like them. The fifth and sixth graders, assumed to attend to all of the presented variability, find experience with 20- and 40-turn figures little help in processing high variability; after exposure to variability well above their assumed processing limit, no change in preference occurs. Only the group of fifth and sixth graders given experience with 10-turn shapes—a degree of variability near their upper limit of processing ability—showed signs of increased preference for 40-turn figures.

The relative simplicity of the foregoing account is inadequate to comprehend the response of the groups of different age and different experience to figures of 5 turns. It will be remembered that no substantial change in preference for figures of low variability was predicted for any group of subjects. Figure 6, on the contrary, shows that the largest obtained dispersion among age groups was in response to 5-turn figures. Why should experience with 5-turn figures lead to an increase in preference for similar figures among children and a decrease among adults? Equally puzzling, why should experience with high variability (40-turn figures) lead to an increase in preference for 5-turn figures among adults and no change or a decrease among children? To the first question, one may reasonably answer that 5-turn figures are so much below the processing limits of adults that exposure to 224 instances of low variability surfeits and bores them. It is less obvious to conclude that children who have looked at so many low-variability figures still find enough uncertainty in 5-turn shapes to interest them. The second question requires an explanation for the paradoxical increase in adult preference for 5-turn figures after 40-turn experience. Clearly, this shift is not merely a turning away from an overload of variability; these same subjects show an increased preference for 40-turn figures as well (Figure 9). But it should be noted again that the increased preference for 5-turn figures shown by adults immediately after experience with 40-turn figures does not appear 24 hours later and that the preference for 40-turn figures is greater 24 hours after experience with similar figures of high variability than it is immediately after exposure. The influence of temporal variation on these phenomena of changed preference for stimulus variability requires further empirical analysis.

SUMMARY AND DISCUSSION

A study of the ability of subjects of three age levels to estimate the amount of stimulus variability contained in random figures ((Study I) supported the postulation of a limit on the processing ability of human beings. Although variation with age was shown, in general all subjects showed decreasing estimation accuracy for figures varying from 5–10 independent turns and a very low but unchanging estimation accuracy for figures of more than 10 turns. This finding accords well with data obtained earlier on preference for random shapes (Munsinger & Kessen, 1964).

Although a postulate of limited processing ability comprehends the data on estimation accuracy and preferences for figures of between 5 and 10 independent turns, it is inadequate for an explanation of the preference response to figures of more than 10 turns. We have shown in previous research that young children prefer figures of high variability more than do adults. Studies II and III of the present report were designed to examine the suggestion that this apparently paradoxical finding was due to the tendency of young children to attend to only part of the variability presented in random figures of many turns. The prediction, tested in Study II, that young children would have differential difficulty in attaching class names to figures of high variability was confirmed. The prediction, tested in Study III, that young children would have differential difficulty in sorting figures of high variability appropriately was also confirmed. Taking these results together with earlier findings (Munsinger et al., 1964), it is our tentative conclusion that young children prefer high variability, and thus appear to be able to process high variability.

largely because they select from figures of many turns that amount of variability which they can handle.

A study of the effect of four levels of experience with variability on the preference behavior of subjects of three ages (Study IV) was designed to test the assumption that change in cognitive structure (and consequent change in preference) would be most likely when subjects were presented with levels of variability just beyond their ability to process. The results obtained were congruent with this postulate only in part. The change in preference response to figures of high variability was generally as predicted; the change in response to figures of intermediate and especially to figures of low variability indicated that revisions of the simple assumption about the nature of cognitive change would be necessary. Specifically, the results of Study IV suggested inter alia that (*a*) children remain responsive to the few-turn figures even though their expressed preference for these figures is initially low; and (*b*) the response of adults (and probably of children) to experience with variability is dependent in significant part on the temporal relations of experience and test of preference.

REFERENCES

ATTNEAVE, F. Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology*, 1957, **53**, 221–227.

ATTNEAVE, F. *Applications of information theory to psychology.* New York: Holt, 1959.

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychological Bulletin*, 1956, **53**, 452–471.

BARTLETT, F. C. *Remembering: An experimental and social study.* London: Cambridge Univer. Press, 1932.

BERLYNE, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

BUTLER, R. A. Discrimination learning by rhesus monkeys to visual-exploration motivation. *Journal of Comparative and Physiological Psychology*, 1953, **46**, 95–98.

GARNER, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HUNT, J. McV. Motivation inherent in information processing and action. In O. J. Harvey (Ed.), *Motivation and social organization: Cognitive determinants.* New York: Ronald, 1963. Pp. 35–94.

KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMANN, J. The discrimination of visual number. *American Journal of Psychology*, 1949, **62**, 498–525.

MANDLER, G. From association to structure. *Psychological Review*, 1962, **69**, 415–427.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior.* New York: Holt, 1960.

MUNSINGER, H., & KESSEN, W. Uncertainty, structure, and preference. *Psychological Monographs*, 1964, **78**(9, Whole No. 586).

MUNSINGER, H., KESSEN, W., & KESSEN, M. L. Age and uncertainty: Developmental variations in preference for variability. *Journal of Experimental Child Psychology*, 1964, **1**, 1–15.

NEWELL, A., SHAW, J. C., & SIMON, H. A. Elements of a theory of human problem solving. *Psychological Review*, 1958, **65**, 151–166.

PIAGET, J. *The psychology of intelligence.* (Trans. by M. Piercy & D. E. Berlyne) New York: Harcourt, Brace, 1950.

PIAGET, J. *The origins of intelligence in children.* New York: International Universities Press, 1952.

ROSS, R. T. Optimum orders for the presentation of pairs in the method of paired comparisons. *Journal of Educational Psychology*, 1934, **25**, 375–382.

SHANNON, C. E., & WEAVER, W. *The mathematical theory of communications.* Urbana: Univer. Illinois Press, 1949.

(Received November 25, 1964)

Psychological Review
1966, Vol. 73, No. 2, 179–192

TOWARD A THEORY OF BEHAVIORAL CONTAGION[1]

LADD WHEELER[2]

Naval Medical Research Institute, Bethesda, Maryland

Behavioral contagion is defined operationally and is contrasted with conformity, imitation, social pressures, and social facilitation. Experiments dealing with contagion are reviewed and theoretical statements derived from this review. The basic theoretical argument is that the lowering of the avoidance gradient in an approach-avoidance conflict is essential to the occurrence of contagion.

Redl (1949), the social scientist most closely associated with the concept of contagion, lists four points as necessary for its occurrence:

1. Existence of an acute conflict area within the imitator; strong impulse urge toward fulfillment of a vehement need on one hand, sufficient pressure from ego or superego forces to keep it down on the other.
2. High degree of lability of the imitator's "personality balance" in the area concerned; impulses strong enough to press for release, controls only just strong enough to prevent that release.
3. Existence of a similar type of strong urge toward impulse expression along the same line in the initiator; his urge must lie in the same direction as that of his imitator.
4. Open acting out in favor of impulse satisfaction by the initiator, with equally open display of an entire lack of fear or guilt.

Combining these four points does not provide a definition of contagion, as the points more nearly constitute a theory than a definition. When Redl and his associates began systematic research on contagion, they defined the concept as follows:

> An incident of behavioral contagion, for the purposes of the present investigation becomes, then, *an event in which a recipient's behavior has changed to become "more like" that of the actor or initiator. This change has occurred in a social interaction in which the actor has not communicated intent to evoke such a change* [Polansky, Lippitt, & Redl, 1950, p. 322].

Unfortunately, the above definition does not adequately distinguish contagion from other types of social influence. For example, the Asch situation fits the definition.

Behavioral contagion[3] will be defined in this paper through the use of reduction sentences (Carnap, 1936,

---

[1] From Bureau of Medicine and Surgery, Navy Department, Research Task MR005.12-2005.01, Subtask 1. The opinions and statements contained herein are the private ones of the writer and are not to be construed as official or reflecting the views of the Navy Department or the Naval Service at large.

[2] I thank Irving Altman and Albert Bandura for their comments on earlier versions of this paper. An extensive correspondence with J. Barnard Gilmore and a careful study of his (1965) paper have been most valuable. His paper may be obtained from him at the University of Waterloo.

[3] The term "behavioral contagion" is used advisedly. Many psychologists, because of certain aspects in the development of the concept of contagion, associate it with the spread of an emotional climate. While emotions do engender like emotions under certain circumstances, this paper will be concerned with overt behavior which may or may not follow or be accompanied by the spread of emotions.

1937). One advantage in using reduction sentences to form a definition is that the definition is then "open." Other reduction sentences may be added to the definition as knowledge of the phenomenon increases. One reduction sentence will be used at this time to define contagion. It is labeled below as S (1).

S (1). If the set of test conditions T_1 exists, then contagion has occurred if and only if Person X performs Behavior N (B_N) where T_1 is specified as follows: (*a*) A set of operations has been performed on Person X which is known to produce instigation toward B_N in members of the class to which X belongs; (*b*) B_N exists in the response repertoire of X, and there are no physical restraints or barriers to prevent the performance of B_N; (*c*) X is not performing B_N; (*d*) X observes the performance of B_N by Person Y.[4]

S (1) defines contagion completely in terms of observable events and conditions. It is necessary to the next section of this paper to draw certain inferences about the processes underlying S (1). If Person X is instigated to B_N, is not physically restrained from performing B_N, and is not performing B_N, the implication is that X possesses internal restraints against the performance of B_N. Thus, there is an intrapersonal conflict between the instigation to perform B_N and the internal restraints against performance of B_N. In other words, X is experiencing an approach-avoidance conflict. S (1) also implies that X's observation of the performance of B_N by Person Y changes the relative strengths of the approach tendency and the avoidance tendency. The implication is supplied by the fact that X changes his behavior after observing the behavior of Person Y. The discussion of what changes take place in the approach and avoidance tendencies will be deferred until the theoretical section of the paper.

An example of the general situation described in S (1) might be appropriate at this point. Every Sunday an individual, X, is frustrated in his quest for sleep by the sound of the power mower of the neighbor across the street, Z. Because of the strong restraints against directing aggression toward a neighbor in our culture, S has confined himself to politely requesting every Sunday that Z wait until afternoon to mow his lawn. One Sunday, as X is approaching Z to make this request, another neighbor, Y, also approaches Z. Y curses Z and threatens him with physical destruction. Without even waiting for Z's reaction to this attack, X also becomes quite aggressive toward Z.

In the remainder of this paper, Person X of both the definition S(1) and the example above will be referred to as the observer. Person Y of S (1) and the above example will be referred to as the model.

RELATIONSHIP BETWEEN CONTAGION AND OTHER TYPES OF SOCIAL INFLUENCE

One may ask why the concept of behavioral contagion is necessary when psychologists already have conformity, imitation, social facilitation, and social pressures (not to mention half a dozen other related terms). Let us briefly examine some of these terms.

The term "conformity" has been used so indiscriminately that one must be somewhat arbitrary in defining it. The mass of recent research on conformity has been concerned with the Asch effect or modifications thereof. Consequently, the definition of conformity presented here will be of Asch-type conformity. Such a definition might be written as follows: If Line *ab* is sufficiently longer than Line *cd* so that Person X has in the past normally given the correct judgment of relative line length when making this judgment in the absence of other individuals, and if Person X is asked to make this judgment immediately after and in the presence of three or more other individuals who have audibly stated that Line *cd* is longer than Line *ab*, then Person X is conforming if and only if he states that Line *cd* is longer than Line *ab*.

---

[4] B_N is not to be taken as a completely specific behavior, but as a behavior class. For example, aggression is a class of behavior, while the verbal statement, "You idiot!," is simply one member of that class. Thus, if the model says, "You idiot!," and the observer says, "I'll poke you in the nose!," both statements are to be considered as B_N.

This definition could, of course, be written in somewhat more general terms. Asch-type conformity is not restricted to the judgment of line length and has, of course, been extended into opinions. But regardless of what is being judged, it is apparent from the above definition that conflict is produced in Person X by the judgments of other individuals. The conflict is between making the subjectively veridical but socially rejected judgment and making the subjectively counterfactual but socially accepted judgment. According to the definition, Person X must resolve this conflict by himself. That is, the actions of the other individuals create the conflict and do not aid in resolution of the conflict.

On the other hand, the reduction sentence, S (1), used to define contagion implies that conflict in Person X exists *prior* to the appearance of other individuals and that the presence of other individuals contributes to conflict resolution. Thus, there is a definite and major difference between conformity and contagion.

An interesting example of the difference between conformity and contagion may be seen in the effect of a nonunanimous majority in the Asch situation. Prior to the appearance of a partner who gives the veridical judgment, the subject (*S*) is instigated to give the veridical judgment, but experiences internal restraints against doing so because the unanimous majority has given a different judgment. *S* resolves this conflict created by the majority by giving the nonveridical judgment; the socially created restraints against giving the veridical judgment are too great. A partner who then begins to give the veridical judgment reduces *S*'s restraints, which is to say that contagion occurs, and *S* ceases to conform to majority opinion. Thus, we have a well-documented situation in which contagion is not only directly opposed to conformity, but appears to be the more powerful of the two phenomena.

At this point in the paper, it has been argued that while contagion and conformity both involve conflict, they differ sharply in the role other individuals play in the conflict. A third term which has been of great importance in research on social influence is "pressures toward uniformity" (Festinger, 1954). What is the relation of this term to contagion and conformity?

In essentials, there is no difference between conformity and pressures toward uniformity. In both cases, the individual possesses an opinion or belief which he then finds to be in conflict with the opinions or beliefs of others. In both cases, the individual must resolve this conflict. Any appearance of differences between the two types of social influence is due to the greater elegance of the pressures toward uniformity formulation and to different research techniques. That is, experiments conducted to investigate pressures toward uniformity have typically involved face-to-face discussions of opinions in which pressures toward uniformity may be evidenced by the individual changing his own opinion (the usual evidence for conformity), by the individual changing the opinions of others (impossible in an Asch-type situation), or by the individual rejecting the others as comparison persons (possibly and likely in an Asch-type situation, but the relevant data are never collected).

Another type of social influence is "social facilitation." According to Thorpe (1956, p. 120), social facilitation is said to occur when "the performance of a more or less instinctive pattern of behavior by one member of a species will tend to act as a releaser for the same behavior in another or in others, and so initiate the same lines of action in the whole group." Thorpe is an ethnologist and would probably forgive students of human behavior for insisting that the behavior need not be instinctive. Bandura and Walters (1963, p. 79) provide the following illustration of social facilitation (or in their terms, of the *eliciting effect* of a model):

> An obvious eliciting effect may be observed in cases in which an adult, who has lost the idioms and pronunciations of the local dialect of the district in which he was raised, returns for a visit to his home. The original speech and pronunciation patterns, which would take a stranger years to acquire, may be quickly reinstated.

One may easily demonstrate social facilitation by reaching for a cigarette during a conversation. Other smokers will also light cigarettes.

Social facilitation was nicely demonstrated by Starch (1911) by having *S*s copy in longhand literary passages written in longhand. Although *S*s were not instructed to do so, they imitated the longhand style in which the passages were written. Starch called this phenomenon "unintentional or unconscious imitation."

In such situations there is no apparent conflict. There would appear to be no definite restraints and no particularly noticeable instigation. By what process then does social facilitation occur? There are several candidate answers which are probably all accurate to some extent.

In the first place, it is probably true that merely conceiving of a behavior increases its probability of occurrence and is usually the first step in performing that behavior. If there are no restraints present, observing and thus conceiving of a behavior is likely to lead step by step to performance. We may call this cognitive-behavioral chaining.

In the second place, there is probably no behavior in one's repertoire which is entirely without restraints. If there are no other restraints, mere inertia constitutes a restraint. Observing someone else perform a behavior reduces these minor restraints; after all, it did not seem to hurt the other fellow.

Third, there are many behaviors which are conditioned responses to the conditioned stimulus of another person performing the behavior. Thus, we frequently yawn in the presence of others who are yawning because all members in the group have approximately the same need for oxygen. After this has happened on a number of occasions, one's personal yawn may be elicited solely by the sight of another person yawning.

In short, social facilitation is mediated by several factors, separately or in combination. It is distinguished from conformity, contagion, and social pressures by lack of any marked conflict.

What of the by now almost generic term "imitation"? The work of Miller and Dollard (1941) is, of course, the classic in the field of imitation. The primary argument in their book is that imitation is learned via reward and punishment and that imitation then generalizes across drives, models, and situations. These authors describe two ways of learning imitation: matched-dependent behavior and copying.

> In both cases, the responses are connected to a cue from an independent subject or model. In both, likewise, the punishment of non-matched responses and the reward of matched responses eventually result in conformity of behavior between the leader and the follower. The essential difference between the two processes is that in matched-dependent behavior the imitator responds only to the cue from the leader, while in copying he responds also to cues of sameness and difference produced by stimulation from his own and the model's responses [p. 159].

After learning to imitate, it presumably makes no difference whether learning took place via matched-dependent behavior or copying. Following established principles of stimulus and response generalization, learned imitation will have more or less of an effect in different situations.

Stated somewhat differently, to the extent that Situation A2 is similar to the learning situation, A1 (i.e., the observer has the same drive; the discriminative and background stimuli are the same; the model is the same; the effective responses are the same), any imitation learned in Situation A1 will manifest itself in Situation A2.

It is apparent that if one wishes to define imitation as I have attempted to define other types of social influence, one could not better the definition of English and English (1958), "action that copies the action of another more or less exactly, with or without intent to copy [p. 253]." Thus, imitation may be considered a generic term subsuming contagion, conformity and social pressures, and social facilitation.

In order to make the relationships among the several uniformity processes clearer to the reader, Table 1 describes social pressures and conformity, contagion, and social facilitation in terms of initial conditions for the observer, the model's behavior, the hypothetical processes occurring in the observer, and finally the observer's behavior.

Table 1 performs the function of distinguishing between the various social influence processes. Another necessary scientific function is to synthesize phenomena, and this function is attempted in Figure 1.

TABLE 1

DYNAMICS OF SELECTED INFLUENCE PROCESSES

Stages in influence process	Behavioral contagion	Social pressures and conformity	Social facilitation
Observer's initial conditions	Instigated to B_N. Internal restraints against B_N.	Instigated to B_P. No restraints.	No restraints against B_N or B_P. No instigation to B_N or B_P.
Model's behavior	Model performs B_N.	Model performs B_N.	Model performs B_N.
Hypothetical processes	Reduction of model's restraints against B_N. Fear reduction.	Creation of restraints against B_P. Conflict between B_N and B_P.	Cognitive-behavioral chaining, CS elicits CR, inertia overcome.
Observer's behavior	Observer performs B_N.	Observer performs B_N (or rejects model or induces model to perform B_P).	Observer performs B_N.

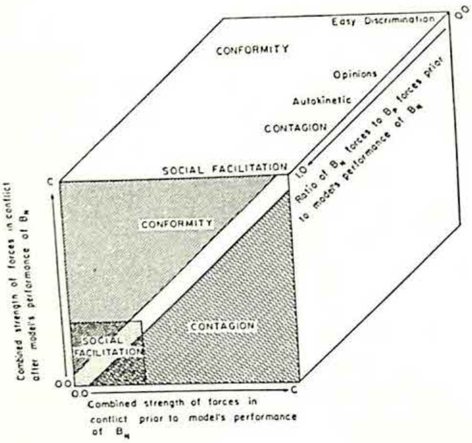
0.0
Easy Discrimination
CONFORMITY
Opinions
Autokinetic
CONTAGION
SOCIAL FACILITATION
1.0
Ratio of B_N forces to B_P forces prior to model's performance of B_N
C
CONFORMITY
CONTAGION
SOCIAL FACILITATION
Combined strength of forces in conflict after model's performance of B_N
0.0
0.0
C
Combined strength of forces in conflict prior to model's performance of B_N

FIG. 1. Three-dimensional model of conformity, contagion, and social facilitation. (The model performs B_N; B_P is any and all behavior or behaviors other than B_N. The strength of forces in the conflict is the sum of the strength of forces in the observer toward B_N and toward B_P. Combined strength of forces ranges from 0.0 to C, where C is the maximum possible in any conflict. The ratio of B_N to B_P forces in the observer is directly related to the probability that B_N will be performed. The case in which the ratio of B_N to B_P forces is greater than 1.0 is not relevant to the discussion.)

The front of the block in Figure 1 is concerned with the combined strength of forces in the observer's conflict both prior to and after the performance by the model of Behavior N (B_N). Note that for any initial level of forces in the conflict, the forces in contagion must be to some degree less after the behavior of the model, and in conformity they must be to some degree greater after the behavior of the model. This may be verified by drawing anywhere on the face of Figure 1 a line parallel to the ordinate. Let the intersection of this line with the abcissa be x, and let the distance along the abcissa from x to the zero point be the *x intercept*. Now draw anywhere on the face of Figure 1 a line parallel to the abcissa. Let the intersection of this line with the ordinate be y, and let the distance along the ordinate from y to the zero point be the *y intercept*. If the two lines so constructed intersect in the contagion area, the *x intercept* will be greater than the *y intercept*. If the two lines intersect in the conformity area, the *x intercept* will be less than the *y intercept*.

If the two constructed lines intersect in the social facilitation area, the *x intercept* may be greater than, less than, or the same as the *y intercept*. This is consistent with the earlier representation of social facilitation as being mediated by some combination of restraint reduction, elicitation of a conditioned response, and cognitive-behavioral chaining. Restraint reduction would, of course, reduce the combined strength of forces, while elicitation of a conditioned response and cognitive-behavioral chaining would act as additional forces and would thus increase the combined strength of forces. Furthermore, social facilitation is shown to involve small forces both before and after the behavior of the model. Social facilitation overlaps with both conformity and contagion, thus describing the twilight zone in which these latter two phenomena may be indistinguishable.

The third dimension of the block in Figure 1 represents the ratio of B_N forces (B_N is the behavior performed by the model) to B_P forces (B_P is any other behavior) prior to the model's performance of B_N. It is indicated that for contagion the B_N forces are slightly less than the B_P forces, and we know from the face of the block that the sum of the B_N and B_P forces is relatively large prior to the behavior of the model. On the other hand, the forces involved in social facilitation are approximately equal, and it is shown on the face of the block that the sum of these forces is small; that is, there is high lability, but no marked conflict. In the case of conformity the B_P forces are greater than the B_N forces, but the ratio varies with the type of situation. In the easy discrimination kind of situation originally used by Asch (1956), B_P forces are far greater than B_N forces. When opinions are at issue (cf. Festinger, 1954), the ratio of B_N to B_P forces is somewhat greater, and the ratio is even greater in the autokinetic situation (Sherif, 1936). This distinction of various conformity situations is, of course, consistent with the well-known fact that conformity varies directly with the ambiguity of the stimuli to be judged.

It should be noted that Figure 1 does not contain the situation in which the ratio of B_N to B_P forces is greater than 1.0 prior to the model's performance of B_N. This is because such a ratio would lead the observer to perform B_N independent of the model's performance of B_N. There would be no social influence.

CONTAGION: EMPIRICAL REVIEW AND THEORETICAL IMPLICATIONS

Little research relevant to behavioral contagion has been conducted. What research that has been done will be organized according to the major question(s) to which the research contributed. After research relevant to any one question has been described, theoretical implications of the research will be discussed. The central theoretical statement to the following discussion is that *behavioral contagion as defined in S (1) is mediated by the lowering of the observer's avoidance gradient in an approach-avoidance conflict.*

Specificity of Response Matching

In an experiment by Grosser, Polansky, and Lippitt (1951), *S*s were told that certain unspecified toys in a collection of toys were forbidden. In this way restraints against playing with any of the toys were created. *S*s played with the toys more when they observed an experimental confederate playing with the toys than when they did not. In addition, *S*s tended to play with toys they had initially rated as most attractive, regardless of which toys the confederate played with. The investigators attributed the successful contagion to a change in the overall perceived safety of the situation.

Similar results with older *S*s were obtained by Wheeler, Smith, and Murphy (1964). Army recruits were placed in a waiting room containing magazines, darts, and shuffleboard. They were told that they could read magazines while they waited. Thus, mild restraints against playing with the games were created. *S*s played with the games more frequently when they observed an experimental confederate playing with one of the games than when the confederate read or simply walked about the room or when the confederate was not in the room. This was true whether or not the quantity of game equipment allowed *S* to play the same game as the confederate. Thus, contagion was demonstrated under conditions that made specific response matching impossible.

These experiments suggest that *depending upon the initial strength of the approach and avoidance gradients, the observer's behavior may or may not be an exact imitation of the model's behavior.* In a situation involving several potential actions directed toward the same end or expressing the same need pattern, observation of the model's performance of Behavior A will reduce the observer's fear of performing A^1 as well as A. Depending then upon the relative strengths of the approach gradients for A and A^1, the observer may perform A^1 as a consequence of witnessing the performance of A.

Balancing of Approach and Avoidance Tendencies

Kimbrell and Blake (1958) created two degrees of thirst in *S*s. With an experimental confederate, *S*s were placed in the vicinity of a water fountain with a sign which read: "Do not use this fountain." The measure of contagion in each of the two thirst conditions was the difference in the percentage of *S*s who drank if the confederate did not drink and the percentage who drank if the confederate did drink. Contagion was greater in the Weak Thirst condition because most *S*s in the Strong Thirst condition drank regardless of what the confederate did.

Freed, Chandler, Blake, and Mouton (1955) obtained somewhat similar results. *S*s approaching the main door of a university building encountered a sign urging use of an inconvenient side door. In three separate conditions, the sign implied three different degrees of prohibition. An experimental confederate either did or did not conform to the injunction in each of the conditions. The measure of contagion in each condition was the difference between the percentage of *S*s obeying the sign when the confederate conformed and the percentage obeying the sign when the confederate deviated. This difference was greatest when the sign implied the least degree of prohibition. When the sign implied strong prohibition, *S*s obeyed the sign whether the confederate did or not.

An implication of these experiments is that *the probability of the occurrence of behavioral contagion is greatest when the avoidance gradient is just slightly higher than the approach gradient.* If the restraints are barely sufficient to prevent overt expression of the approach tendency, a small reduction in fear will be sufficient to produce the approach behavior. If the restraints are far stronger than the approach tendency, massive fear reduction is necessary to produce a change in behavior. If the avoidance gradient is below the approach gradient, the behavior will be performed whether the model is present or not.

Consequences to the Model of his Behavior

Bandura, Ross, and Ross (1963b) exposed children to a television program depicting an aggressive model. In one condition, the model was rewarded for his aggression; that is, the model obtained good things for being aggressive. In a second condition, the model was thrashed for this aggression. After viewing the program, children were placed in a room containing numerous toys that could be used to reproduce the model's aggressive response patterns. Children who had observed the rewarded model were more aggressive than children who had observed the punished model.

Walters, Leat, and Mezei (1963) placed children in the vicinity of an attractive assortment of toys with which they were forbidden to play. The children were then shown a film of another child playing with the toys. Some of the children saw the child rebuked by his mother for playing with the toys. Others saw the child rewarded through his mother's nurturant interaction. After viewing the film, the children were left alone in the room containing the toys. Children who had seen the model rewarded played with the toys more frequently than children who had seen the model punished.

It is apparently not necessary for the observer to know the consequences of the model's behavior, however. Bandura, Ross, and Ross (1963a), Walters and Thomas (1963), and Walters, Thomas, and Acker (1962) found *S*s to be more aggressive after watching an aggressive model who had not been obviously rewarded or punished for aggression. It is probably the case that if the observer expects the model to be punished, the lack of any overt punishment is perceived as rewarding.

Walters and Parke (1964) and Walters, Parke, and Cane (1965) found that children violated prohibitions against playing with toys less after having observed a punished model than after having observed a rewarded model or a model who was neither rewarded nor punished.

An implication of these studies is that *the observer vicariously performs the behavior performed by the model, and the observer vicariously experiences the consequences of the behavior which are overtly experienced by the model. To the extent then that the model is rewarded or not punished, the observer's avoidance gradient is lowered.*

Characteristics of Effective Models

Polansky et al. (1950) and Lippitt, Polansky, and Rosen (1952) found that children high in prestige were the most effective models for contagion. However, many of the incidents of contagion tabulated and analyzed did not involve restraints, and thus the conclusions from the two studies are not as relevant to the current discussion as they might be. A field experiment on the contagion of aggressive behavior, done within the larger study by Polansky et al. (1950), found effective models to be highly impulsive. This may mean simply that impulsive boys acted sooner than others and thus had more opportunity to become models.

Lefkowitz, Blake, and Mouton (1955) investigated the contagion of pedestrians crossing against a red light. Contagion was greater when the initiator was a well-dressed, high-status individual than when he was poorly dressed.

The results with regard to the status of the model suggests that general characteristics of the model may have the same effect as immediate consequences of the model's behavior. If the observer perceives the model engaging in behavior which has no immediate visible consequences, the observer may estimate the likelihood of the model's subsequent punishment or reward. One basis for such an estimate is the previous success of the model in obtaining rewards and avoiding punishments. Thus, *to the extent that the model appears to be or is known to be generally rewarded or not punished for whatever behavior he emits, the observer's avoidance gradient is lowered for whatever behavior the model performs.*

Characteristics of the Observer

Polansky et al. (1950) and Lippitt et al. (1952) found that observers most susceptible to influence through contagion had high prestige. The field experiment on the contagion of aggression, also reported by Polansky et al. (1950), indicated that impulsive children were most likely to be influenced toward aggression through the observation of aggressive models. These findings duplicated exactly those concerning effective models.

Smith, Murphy, and Wheeler (1964) found no relationship between conforming in an Asch-type situation and being influenced by a model to violate a prohibition against playing with games available in a waiting room. Moreover, while the usual relationship between conformity and authoritarianism was obtained, there was no relationship between contagion and authoritarianism.

There is evidence that susceptibility to contagion is increased by unusual or extreme physiological states. Novakovsky (1924) linked starvation, fatigue, illness, and extreme cold to susceptibility to modeling influences. A number of the incidents mentioned involved behavior that possessed potentially negative consequences. Schachter and Singer (1962) found that an adrenalin injection increased the contagion of anger, particularly if *S* could not explain his unusual physiological state as a product of the injection.

Bandura et al. (1963b) investigated sex differences in observers in an experiment on the contagion of aggression. *S*s watched either a televised nonaggressive model, a model punished for his aggression, or a model rewarded for his aggression. After observing the adult male models, the children were placed in a playroom, and aggressive behavior was observed. Compared to a control condition in which no model was observed, total aggression expressed by the females was lowered by the nonaggressive model and not affected by the other models. In contrast, the total amount of aggression produced by the males was decreased in the model-punished condition and increased in both the model-rewarded condition and the condition in which a nonaggressive model was observed. These results were explained in a reasonable although post hoc manner, and it is not necessary to repeat the explanation here.[5]

The evidence relevant to characteristics of the observer having been reviewed, no general statement tying the experiments together is readily apparent. We will let the review stand on its own merits. The only point that might be stressed is that the kind of person most susceptible to influence through contagion does not appear to be anxious, authoritarian, or of low self-esteem.

Locus of Restraints

It has been argued in this paper that contagion is mediated by restraint reduction or fear reduction. Thus, it may be useful in making predictions concerning contagion to recognize that the locus of restraints may be different from one situation to another. As the locus of restraints changes, what is feared should change. In turn, situational determinants of contagion may change. While there has been no research directed toward this possibility, certain predictions present themselves.

If the restraints are group derived, the observer will fear rejection by the group for his lack of impulse control. The most effective model should be a solid supporter of the group norms, an individual who is quite unlikely to be rejected. If this individual performs the forbidden action, he is in essence declaring that the group no longer forbids the action, thus reducing the observer's fear. However, if there is too much of a status differential between the model and the observer, the observer may feel that the model is using "idiosyncrasy credits" (Hollander, 1958) which he himself does not possess. In such a situation, fear would not be reduced, and contagion would not occur. To the extent that the observer is attracted to the group and does not have access to other desirable groups, fear of rejection should be strong, and contagion should be less likely to occur. To the extent that other group members express disapproval of the model's action, contagion will not occur. This effect will be most striking when disapproval is expressed by high-status group members. In general, when the restraints are group derived and a group member performs the forbidden act, the behavior should spread quickly throughout the group. Every member that performs the act reduces the restraints further.

---

[5] Other studies have shown the sex of the observer to influence contagion. Such results are important to the present discussion in pointing out that socialization creates restraints and dispositions that affect susceptibility to contagion in different ways depending upon what behavior is involved in the contagion.

In a different situation, the locus of restraints is the observer's superego. The observer fears guilt and regret for engaging in the behavior. Thus, the model's action must somehow reduce the anticipation of guilt by the observer which is possible by making the action seem more "right" and less "wrong." For example, Bandura and Walters (1959) have suggested that adolescents are more likely to engage in sexual intercourse during double dating than at other times. When there are superego restraints, the apparent morality of the model should be of great importance. An adolescent girl is not likely to have her restraints against sexual behavior lowered if the model is either a prostitute or a nun. The prostitute model could not reduce the observer's fear of guilt because of the model's apparent lack of sexual morality. The nun model would be rejected as a false nun. The most effective model should have slightly higher apparent sexual morality than the observer. Another prediction is that to the extent that the model performs the behavior without overt guilt or fear responses, the observer's fear should be reduced.

In a still different situation, the locus of restraints is an authority figure. The optimal conditions for eliciting contagion in such a situation are complex. As pointed out earlier, the most effective model is often one who generally appears to escape punishment for his behavior. But the model who fits this description perfectly is often not an effective model because he is not subject to the authoritarian restrictions in the first place. Thus, a speeding police car is not likely to reduce restraints against breaking the speed limit.

The most effective model for reducing authoritarian restraints would be one who performed the behavior in the presence of the authority without punishment by the authority. In addition, the model must originally appear to be subject to the same authoritarian restriction as the observer.

An experiment by Milgram (1965) provides an example of contagion when the locus of restraints is an authority figure. *S*s were instigated to cease shocking another individual, but they were ordered by an authority to continue the shocking. Two models who were apparently under the same orders from the authority refused to continue the experiment and were not punished by the authority for their disobedience. Virtually all *S*s ceased to obey the authority. On the other hand, models who obeyed the authority had no effect at all, as one would predict from the theory of contagion.

Unlike the situation in which the locus of restraints is the superego, similarity between the observer and model is probably not relevant. This is because the observer fears punishment from an external source, and any model who performs the behavior without negative consequences will reduce restraints, no matter how dissimilar he may be to the observer.

There is no intention to imply that the discussion of group-derived restraints, superego restraints, and authoritarian restraints is exhaustive; there are undoubtedly other common loci of restraints. Nor is there any intention to imply that the usual situation has only *one* locus of restraints. The intention is merely to suggest the potential predictive value of specifying the locus of restraints.

Interpersonal Consequences of Contagion

Festinger, Pepitone, and Newcomb (1952) hypothesized that when members of a group are deindividuated, internal restraints are reduced, and previously suppressed behaviors are performed. This in turn leads to satisfaction with the group. Essentially the same hypothesis was suggested by Hoffer (1951). Some support for this hypothesis was provided by Festinger et al. (1952) who found (*a*) a negative correlation between the frequency of negative statements about parents and the ability to remember which person in the group said what, and (*b*) a positive correlation between the frequency of negative statements about parents and the desire to have further discussions with the other group members. The first correlation was interpreted as a positive relationship between restraint reduction and deindividuation, while the second correlation was interpreted as a positive relation between restraint reduction and attraction to the group.

Singer, Brush, and Lublin (1964) found that in groups of girls discussing pornographic literature, members were more attracted to groups in which obscene words were spoken. This was interpreted as a positive relation between restraint reduction and attraction to the group.

Deindividuation is a feeling of anonymity, and one would expect it to exist to the extent that the individual cannot be singled out as the only person performing a behavior. Thus, observation of a model performing a behavior should produce in the observer some feeling of anonymity with regard to his own performance of that behavior. If the model appears to be "getting away with it," the overall result should be a lowering of the observer's restraints against performing the behavior, the restraints being lowered both by the feeling of anonymity and by the knowledge that the behavior will not be punished. Now it is probably true that whenever restraints are reduced, by whatever means, to the point of behavior occurring and going unpunished, there will be satisfaction with the situation. One type of satisfaction with the situation would be "liking" for the model and a knowledge that he and the model have something in common. It has been shown in a series of studies that when an individual perceives some attribute in common between himself and a model, he will tend to perceive himself as having other of the model's attributes (Burnstein, Stotland, & Zander, 1961; Stotland & Dunn, 1962, 1963; Stotland & Hillmer, 1962; Stotland & Patchen, 1961; Stotland, Zander, & Natsoulas, 1961). From these considerations, we may predict that *after the occurrence of contagion in a situation in which the observer is not punished for his behavior, the observer will like the model more than previously and will feel more similar to him on a variety of dimensions, that is, he will be less individuated from the model.*

Note that while in the Festinger et al. (1952) formulation deindividuation leads to restrain reduction, the contagion formulation suggests that most of the deindividuation occurs after the restraint reduction.

CONCLUDING REMARKS

This paper has focused upon examples of contagion which involve socially undesirable behavior, such as aggression, adolescent sexual intercourse, and violation of formal prohibitions.

But it is not necessary that contagion be confined to socially undesirable behavior. Experiments have shown that volunteering behavior is increased by the observation of others volunteering (Rosenbaum & Blake, 1955; Schachter & Hall, 1952). The latter authors specifically attributed the effect of volunteering models to the reduction of group-derived restraints.

The incident in the Kew Gardens section of New York in which a number of people observed a woman attacked three times and eventually murdered could probably have been prevented by one individual. Had any one individual made a visible attempt to aid the victim, the fear would probably have been reduced in other of the observers, who would have then acted.

This paper has dealt with socially undesirable behavior because when the behavior elicited through contagion is socially desirable, it is sometimes difficult to rule out other explanations. For example, there are probably strong pressures upon an individual to volunteer when those around him are volunteering. However, it should be emphasized that *theoretically* contagion is not limited to socially undesirable behavior. The theory is concerned with all behavior which the observer perceives as possibly bringing negative consequences to the performer.

REFERENCES

ASCH, S. E. Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs,* 1956, **70**(9, Whole No. 416).

BANDURA, A., ROSS, DOROTHEA, & ROSS, SHIELA A. Imitation of film-mediated aggressive models. *Journal of Abnormal and Social Psychology,* 1963, **66**, 3–11. (a)

BANDURA, A., ROSS, DOROTHEA, & ROSS, SHIELA A. Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology,* 1963, **66**, 601–607. (b)

BANDURA, A., & WALTERS, R. H. *Adolescent aggression.* New York: Ronald, 1959.

BANDURA, A., & WALTERS, R. H. *Social learning and personality development.* New York: Holt, Rinehart & Winston, 1963.

BURNSTEIN, E., STOTLAND, E., & ZANDER, A. Similarity to a model and self-evaluation. *Journal of Abnormal and Social Psychology,* 1961, **62**, 257–264.

CARNAP, R. Testability and meaning: I. *Philosophy of Science,* 1936, **3**, 420–471.

CARNAP, R. Testability and meaning: II. *Philosophy of Science,* 1937, **4**, 1–40.

ENGLISH, H. B., & ENGLISH, AVA C. *A comprehensive dictionary of psychological and psychoanalytical terms.* New York: Longmans, Green, 1958.

FESTINGER, L. A theory of social comparison processes. *Human Relations,* 1954, **7**, 117–140.

FESTINGER, L., PEPITONE, A., & NEWCOMB, T. Some consequences of deindividuation in a group. *Journal of Abnormal and Social Psychology,* 1952, **47**, 382–389.

FREED, A., CHANDLER, P. J., BLAKE, R. R., & MOUTON, JANE S. Stimulus and background factors in sign violation. *Journal of Personality,* 1955, **23**, 499.

GILMORE, J. B. When should we predict imitative behavior; when shall we understand it? Paper read at Society for Research in Child Development, Minneapolis, 1965.

GROSSER, D., POLANSKY, N., & LIPPITT, R. A laboratory study of behavioral contagion. *Human Relations,* 1951, **4**, 115–142.

HOFFER, E. *The true believer.* New York: Harper, 1951.

HOLLANDER, E. P. Conformity, status, and idiosyncrasy credit. *Psychological Review,* 1958, **65**, 117–127.

KIMBRELL, D., & BLAKE, R. R. Motivation factors in the violation of a prohibition. *Journal of Abnormal and Social Psychology,* 1958, **56**, 132–133.

LEFKOWITZ, M. M., BLAKE, R. R., & MOUTON, JANE S. Status factors in pedestrian violation of traffic signals. *Journal of Abnormal and Social Psychology,* 1955, **51**, 704–706.

LIPPITT, R., POLANSKY, N., & ROSEN, S. The dynamics of power. *Human Relations,* 1952, **5**, 37–64.

MILGRAM, S. Liberating effects of group pressure. *Journal of Personality and Social Psychology,* 1965, **2**, 127–134.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

NOVAKOVSKY, S. Arctic or Siberian hysteria. *Ecology,* 1924, **2**(5).

POLANSKY, N., LIPPITT, R., & REDL, F. An investigation of behavioral contagion in groups. *Human Relations,* 1950, **3**, 319–348.

REDL, F. The phenomenon of contagion and "shock effect" in group therapy. In K. R. Eissler (Ed.), *Searchlights on delinquency.* New York: International Universities Press, 1949. Pp. 315–328.

ROSENBAUM, M., & BLAKE, R. R. Volunteering as a function of field structure. *Journal of Abnormal and Social Psychology,* 1955, **50**, 193–196.

SCHACHTER, S., & HALL, R. Group-derived restraints and audience persuasion. *Human Relations,* 1952, **5**, 397–406.

SCHACHTER, S., & SINGER, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, **69**, 379–399.

SHERIF, M. *The psychology of social norms.* New York: Harper, 1936.

SINGER, J. E., BRUSH, CLAUDIA A., & LUBLIN, SHIRLEY C. Conformity and deindividuation: Studies in conflict and restraint. Research Bulletin No. 44, June 1964, Pennsylvania State University.

SMITH, S., MURPHY, D. B., & WHEELER, L. Relation of intelligence and authoritarianism to behavioral contagion and conformity. *Psychological Reports,* 1964, **14**, 248.

STARCH, D. Unconscious imitation in handwriting. *Psychological Review,* 1911, **18**, 223–228.

STOTLAND, E., & DUNN, R. E. Identification, "oppositeness," authoritarianism, self-esteem, and birth order. *Psychological Monographs,* 1962, **76**(9, Whole No. 528).

STOTLAND, E., & DUNN, R. E. Empathy, self-esteem, and birth-order. *Journal of Abnormal and Social Psychology,* 1963, **66**, 532–540.

STOTLAND, E., & HILLMER, M. L., JR. Identification, authoritarian defensiveness, and self-esteem. *Journal of Abnormal and Social Psychology,* 1962, **64**, 334–342.

STOTLAND, E., & PATCHEN, M. Identification and changes in prejudice and in authoritarianism. *Journal of Abnormal and Social Psychology,* 1961, **62**, 265–274.

STOTLAND, E., ZANDER, A., & NATSOULAS, T. Generalization of interpersonal similarity. *Journal of Abnormal and Social Psychology,* 1961, **62**, 250–256.

THORPE, W. H. *Learning and instinct in animals.* London: Mathuen, 1956.

WALTERS, R. H., LEAT, MARION, & MEZEI, L. Response inhibition and disinhibition through empathetic learning. *Canadian Journal of Psychology,* 1963, **16**, 235–243.

WALTERS, R. H., & PARKE, R. D. Influence of response consequences to a social model on resistance to deviation. *Journal of Experimental Child Psychology,* 1964, **1**, 269–280.

WALTERS, R. H., PARKE, R. D., & CANE, VALERIE A. Timing of punishment and the observation of consequences to others as determinants of response inhibition. *Journal of Experimental Child Psychology,* 1965, **2**, 10–30.

WALTERS, R. H., & THOMAS, E. L. Enhancement of punitiveness by visual and audiovisual displays. *Canadian Journal of Psychology,* 1963, **17**, 244–255.

WALTERS, R. H., THOMAS, E. L., & ACKER, C. W. Enhancement of punitive behavior by audiovisual displays. *Science,* 1962, **136**, 872–873.

WHEELER, L., SMITH, S., & MURPHY, D. B. Behavioral contagion. *Psychological Reports,* 1964, **16**, 159–173.

(Received December 1, 1964)

VOL. 73, NO. 3 MAY 1966

PSYCHOLOGICAL REVIEW

SHORT-TERM VERBAL MEMORY AND LEARNING[1]

LLOYD R. PETERSON

Indiana University

A number of paired-associate experiments are examined for evidence of the interaction of a short-term storage with a long-term learning mechanism. Differential effects depending on length of the retention interval are described for duration of the spacing interval between 2 presentations of individual pairs, duration of presentations, and number of repetitions. The rapidly decreasing initial portion of the retention curve with multiple paired associates is considered to be the result of interference and the passage of time. The later portion of the short-term retention curve is considered to largely reflect an associative mechanism which may produce stronger associations with time. A model of the interaction of these mechanisms with a guessing component is presented.

Hebb (1949) has written a detailed account of a possible relationship between short-term memory and learning. He hypothesized that recall for a brief period of time following stimulation was made possible by the reverberation of neural circuits, and that this reverberation also participated in the formation of a structural memory trace. A single trial by itself was not considered to leave a lasting aftereffect.

More recently, research in short-term verbal retention has provided data from which inferences have been made which appear to support Hebb's concept of a transient activity trace. Brown (1958) obtained results after a single presentation which he interpreted in terms of a rapidly decaying trace. He considers that decay is rapid regardless of whether or not interfering activities occur. Repetition does not strengthen the trace, according to this position, but only puts off the onset of the decay. Broadbent (1963) suggests that it is possible that decay is characteristic of traces which have been repeated several times, as in learning over trials. He holds that short-term memory deteriorates rapidly because of the time that interfering activities take, rather than due to the similarity of the activity to the material being tested. In this respect he considers short-term memory to differ from long-term memory, in which retention has been shown to be related to similarity of the interfering activity to the learning.

In contrast to this point of view there is a growing trend for many investigators to consider that short-term retention involves processes similar to those involved in long-term retention. For instance, Peterson and Peterson (1959) have suggested that a brief presentation of three unrelated letters can be conceptualized as a trial of learning. They tested recall after intervals up to 18 seconds during which the subject engaged in counting activity. Latencies were often long, and their probabilistic nature resembled that of recall in a learning experiment. Further, repetition improved recall in an orderly fashion. Again, Hebb (1961) has modified his position as a result of a verbal memory experiment. He gave subjects a succession of immediate-memory tests in which every third sequence of digits was ordered identically. There was an improvement with repetition even though different orders of the same digits were being tested between repetitions. This suggested to Hebb that something more permanent than an activity trace resulted from a single presentation. Melton (1963) elaborated Hebb's experiment, concluding that a structural trace was produced by a single presentation, and that short-term memory operates by essentially the same mechanisms as does long-term memory. He considered that the effects of various kinds of interference in producing short-term forgetting, as well as the effect of repetitions in improving short-term retention, are further evidence of the continuity of short- and long-term retention.

---

[1] The investigations of the author have been supported by National Science Foundation Grant GB 704. The paper was written at the Institute of Human Learning, University of California, Berkeley. The author is indebted to N. H. Anderson and E. Tulving for criticisms of portions of drafts of the paper.

The present paper outlines a position which accepts the presence of learning mechanisms in short- as well as long-term retention, but which also provides for a recency mechanism which interacts with the learning mechanisms. It is suggested that these processes work concurrently, and that the effective role of an individual process varies with the recency of the event to be remembered, as well as being dependent on various situational variables. Learning is initiated immediately on one presentation, but its presence shortly after presentation is largely masked by the operation of a recency mechanism. Memory for recent events is largely effected by this short-term mechanism, and memory drops off rapidly when events of the same class succeed each other. Memory for events in the more distant past is made possible by learning mechanisms, and it need not decrease over the course of a short interval, but may increase. The relative contribution of each mechanism normally varies as a function of the length of the retention interval and the nature of the events occurring during the interval. Evidence for the manner of operation of these mechanisms will be sought in variation of the composition and duration of intervals separating various events in tests of retention.

Most of the experiments which will be considered deal with paired associates. A paired-associate task would seem to be ideal for the investigation of relationships between short-term memory and learning. It has been studied extensively through the acquisition of lists of pairs and by tests of retention after intervals of time measured in minutes, hours, and days. More recently, studies have been reported in which retention of paired associates has been tested over intervals measured in seconds from the presentation of an individual pair to the cue for its recall. The asymptotes of many such short-term retention curves have been reached at intervals which approximate the time between successive occurrences of individual pairs in list learning over trials. Experimenters using the anticipation technique with lists have taken recall at such an average interval between pairings to be indicative of learning, in contrast to retention which is studied over longer intervals. To the extent that situations can be arranged so that short-term retention studies are similar to list studies in important respects, performance at the asymptote of the short-term retention curve can be considered to reflect the same phenomenon that is reflected by anticipations over trials.

INTERFERENCE

There is considerable evidence to indicate that interference from other activities is a major source of forgetting in short-term tests. However, the means by which its effects are produced require some consideration. Interference comes in various forms, depending on the composition and timing of the activities involved.

Retroactive interference of the classical type in which a new response is paired with an old stimulus has been found in the short-term retention of paired associates (Peterson, Saltzman, Hillner, & Land, 1962). The standard condition consisted of tests of a single pairing after a 10-second interval filled with presentations and tests of other pairs. The retroaction condition differed only by the pairing of the cue word with a new response midway through the retention interval. Recall of the initial pairing was correct in .27 of the tests in the standard condition, while it was only .14 in the retroaction condition.

Evidence for proactive interference of a complementary variety has been reported by Friedman and Gelfand (1964). They successively paired two different responses to two stimuli having a common element. Presentations and tests of other pairs filled the intervals between pairings and between the last pairing and the test for retention. Recall of the second pairing was slightly lower than recall in a control condition in which only one response had been paired with the common stimulus element. The proaction and control conditions were tested equally often at various stages of the session, and the interpretation is clearer than in cases where decrement over stages of a session is explained in terms of proactive interference.

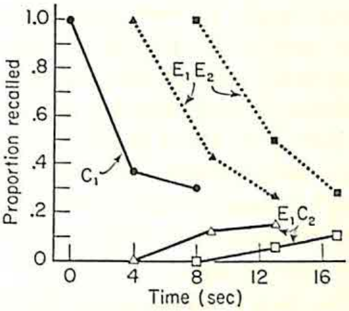

FIG. 1. Recall on a second test in relation to recall on a first test. (C_1 indicates the curve for correct recalls on Test 1. E_1E_2 indicates the dependent probabilities that the same incorrect response occurred on Test 2 as on Test 1. E_1C_2 indicates the dependent probabilities that the correct response occurred on Test 2, given an error on Test 1.)

Data of relevance to the interaction of competing responses to the same cue are provided by an experiment in which two successive tests followed an initial presentation of a pair (Peterson, Saltzman, Hillner, & Land, 1962). The tests did not provide knowledge of results. The pairs were combinations of words and numbers, and the presentation or test of other pairs filled intervals. The data graphed in Figure 1 are from conditions in which either 4 or 8 seconds elapsed between presentation and first test, while another 4 or 8 seconds elapsed between first and second tests. Correct recall on the first test was found to drop off rapidly. Recall on the second test, given an error on the first test, depended on both the retention interval before the first test and the interval from first to second test. Recall was correct in .12 of the cases when the second test followed the first by 4 seconds and .15 when this interval was 8 seconds.[2] Recall in both conditions was significantly higher than the chance level for 15 alternatives ($t = 2.31$, $p < .05$; $t = 3.87$, $p < .001$). In the case of pairs not recalled 8 seconds after presentation, recall did not go appreciably above the chance level on the second test.

There was a suggestion of an overall tendency for recall to improve as time increased since the error on the first test ($F = 4.58$, $df = 1/35$, $p < .05$). It seems possible that responses which were given incorrectly on the first test may have provided an interference effect that decreased with time. Evidence for this interpretation is provided by the conditional probability that an incorrect response on the first test would be followed by the same incorrect response on the second test. These conditional probabilities decreased over the interval following the first test, and the decrease was about the same for both intervals between presentation and first test. The forgetting curves for these errors resemble that of the correct response, suggesting a short-term storage equally efficient for pairings read correctly from the drum as for pairings spoken in error. It seems plausible that an erroneous response made to a cue competed with later correct recall to that same cue. As the erroneous pairing was forgotten, recall of the correct response recovered. The interaction of the initially paired response with the erroneously paired response may provide a paradigm for competition generally. The last occurring event is very strong initially, but it decreases with time to become approximately equal in strength to the earlier event, as both approach the asymptotes of their retention curves. There is a relative increase in strength of the earlier pairing as time passes.

Since pairings which were incorrect on the first test were later correct in substantial numbers on the second test, the results are evidence against any simple all-or-none interpretation of short-term forgetting. A pairing is not forgotten once and for all, but it may become available later. Of course, if two mechanisms are involved in short-term retention, then one could lose information gradually, while the other had all-or-none characteristics. The short-term mechanism could involve gradual cessation of activity, while the learning mechanism produced abrupt changes. There is some suggestion that this might be the case in the finding of slower recovery when an 8-second interval elapsed from presentation to test. Recovery over the intervals tested reached only the level expected by guessing in this condition. It is not possible to judge from the curve whether further recovery might be expected at longer intervals.

NUMBER OF PAIRS

Pairs not having members in common with the pair under test can also be shown to provide competition at recall. The decrease in recall that comes with an increase in the number of pairs recently presented is such an instance. The limiting comparison of differing numbers of pairs arises when retention of one pair is compared with retention of two pairs. Very little forgetting of a single pair of words has been found over intervals up to 16 seconds filled with counting activity, while marked forgetting has been found with two pairs (Peterson & Peterson, 1962). Whether this effect of one additional pair is due to a difference in retention of associative learning is open to question. With only one pair there can be no test of a discrimination between cues. The subject responds with the only response that was presented, and recall in this case is for an individual item. In the case of retention of two pairs, it must be noted that there was a primacy effect. The recall of the first pair was better than recall of the second pair, time since presentation being equal. Since considerations of retroactive and proactive interference would not suggest an inequality in this direction, another factor or factors would appear to be involved. A greater disparity in stimulus conditions from presentation to test in the case of the second pair, when the stimulus situation is considered to extend over recent events, is one consideration. Another is the likelihood that learning of the second pair is lessened by reason of the previous pair having occurred, an effect which may be referred to as associative interference (Goggin, 1964).

---

[2] The values for these conditions listed in the original paper are incorrect.

Comparisons involving from two to five pairs of words have been investigated by Murdock (1963). Recall for one pair of a sequence was tested immediately after the last pair had been presented. Under these circumstances recall was very high in the case of two pairs, and the biggest difference occurred when the number of pairs was increased from two to three. The higher recall in this case for two pairs as contrasted with that found in the experiment of Peterson and Peterson (1962) seems to stem from the fact that Murdock tested immediately after the presentation of the second pair. There is evidence that the last pair of a sequence can be recalled immediately close to 100% of the time; indeed, the subject can tell the experimenter what both members of the pair were without cuing. If, then, two pairs have been presented and the subject is tested immediately by being given one member of the first pair as cue, it seems unlikely that any genuine discrimination is required to give the other member of the pair. The subject has the last pair immediately available, and he sees the third of the four words. The fourth can be given by elimination.

The main interest in Murdock's (1963) experiment lies in the comparisons of three, four, and five pairs. Retention curves were plotted in terms of how many other pairs intervened between presentation of an individual pair and its test. Thus, the case of two intervening pairs could be compared for all three lengths of sequence. However, the presentation of the pair tested would be the first in a sequence of three pairs, while it would be the third in a sequence of five pairs. The slopes of the retention curves over one or two intervening pairs were similar for three, four, or five pairs in the sequence. Retention dropped off rapidly. The chief differences lay in the asymptotes reached, the level being lower the greater the total number of pairs.

It may be hypothesized that the major effect of the competition at recall that varies with number of pairs is on the guessing probability. With more responses available the possibility of guessing correctly is diminished. This point of view is similar to that put forward in connection with list learning with differing numbers of alternatives and length of list held constant. The equation for a learning curve has been fitted to data for one number of alternatives, and performance has been successfully predicted for a list with another number of alternatives without any change in the learning parameter. This was accomplished by taking as the guessing estimate the reciprocal of the number of alternatives (Bower, 1962; Estes, 1961).

Recall failure that occurs due to competition can be thought of as a decreased capacity to discriminate among response alternatives on one basis or another. If the last presentation is to be recalled in the absence of cues, as in the recall of individual items, competition of other recently presented items should be strong, since the discrimination depends on recency alone. Thus, with successive tests performance deteriorates (Keppel & Underwood, 1962). When a basis for discrimination other than recency is provided, as when a number is tested after a series of tests with letters, recall of the item of a new class is as good as recall on the first test of a session (Wickens, Born, & Allen, 1963). In the case of responses for which there are unique cues, competition should be at a minimum, for to the extent that cue-response contingencies have actually been established, competition should be absent. This contrasts with the case where more than one response has been associated with the same cue. Here the cue-response contingency contributes to the competition. In the light of these considerations competition of responses with unique cues can be said to affect the guessing probability, that is, the subject's ability to produce the correct response apart from the influence of the cue.

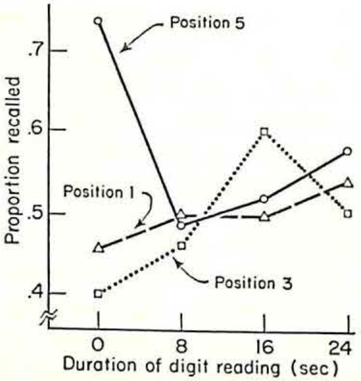

FIG. 2. Recall of a list of five pairs as a function of duration of intervening digit-reading activity.

REMINISCENCE

The activity filling the retention interval in most of the studies mentioned previously has consisted of the presentation of materials of the same class as that testing recall. When intervening activity is used consisting of material of a categorically different kind in a nonlearning task, marked forgetting over short intervals is also found. However, the same type of study has also found evidence of reminiscence with longer durations of this same activity (Peterson, 1966). A list of five pairs of words was presented once, and recall was tested by a random sequence of first members of each pair after from 0–24 seconds. The intervening activity from end of presentation to beginning of test consisted of the reading of random sequences of digits from a memory drum. The retention curves for three of the positions in presentation averaged over all positions in the test sequence are shown in Figure 2. In the case of the fifth position in presentation the digit reading resulted in forgetting over the early portion of the retention curve. This could be attributed to interference from earlier pairs, since proactive interference is considered to increase with time. However, such an interpretation would predict forgetting generally rather than reminiscence. A decreasing and then increasing type of retention curve may be inferred for the first and third positions in presentation. These start lower in the figure than the fifth position, since the early decreasing portion of the retention function can be assumed to have already occurred during the presentation of the last pairs in the list. A decreasing and then increasing curve is also found if the last pair in presentation is observed only on the first test of the sequence. Thus, a nonlearning activity using categorically different materials was found to produce forgetting over short intervals and reminiscence over longer intervals. It would not be expected that recall would continue to improve indefinitely. An experiment has been reported which found forgetting to increase after there had earlier been reminiscence (Keppel & Underwood, 1961).

The rapid forgetting of the last pair could be described as due to proactive interference. This would follow from data previously described showing that without the presentation of the previous pairs there would be little forgetting. Somehow earlier pairs interfere with later pairs, as time passes since the last pairs were presented. The question of why the interference increases with time requires an answer. Let us assume that a recency mechanism loses stored information during any intervening activity that prevents rehearsal. As the subject becomes unable to answer on the basis of the short-term store, competition among recently presented responses enters in to influence his response. If no previous pairs other than the tested pair had been presented, the subject could respond correctly simply by giving the only available response. The total number of responses determines the subject's response because of the failure of the recency mechanism.

Such an explanation by itself fails to account for the reminiscence. Another mechanism would seem to be required, an associative learning mechanism which increases in effectiveness with time. This may be conceptualized as a consolidation of the association. The recall test may be considered to be a disrupting influence which is less severe for a consolidated association.

An alternative explanation in terms of differential forgetting might be possible. The availability of very recent presentations may diminish with time, permitting the cue to have greater weight in determining the subject's responses. There is seen to be a short-term mechanism as well as an associative learning mechanism required in this explanation, also.

The Role of Time

Competition has been noted to play a strong role in forgetting over short intervals, and the question of interference or decay does not permit an answer solely in terms of decay. The term decay will be taken to mean forgetting which would occur no matter how dissimilar preceding and intervening activities were to the tested material. There are a number of experimental results which are difficult for an interference theory to explain. For example, Conrad (1964) has presented letters visually and found confusions in recall which are systematically related to acoustic confusions in auditory presentations under noisy conditions. He interprets the data as showing systematic errors even after the likelihood had been removed that such errors were perceptual. He notes that decay theory demands the existence of partially decayed memory traces. Such traces should yield incorrect memories which are systematically rather than randomly related to the stimuli which were present. His findings are therefore considered by Conrad to be evidence for decay.

Further evidence of forgetting under circumstances that are difficult for an interference theory has been provided by Keppel and Underwood (1962). In their first experiment a brief auditory presentation of a trigram was followed by counting activity during the retention interval. Recall was down to about 50% after 18 seconds on the first test of the session. Keppel and Underwood suggest that there could have been proactive inhibition from experiences prior to the experimental session. However, in view of a finding by Loess (1964) that a previous session 3 weeks earlier with similar materials did not affect recall in the later session, an explanation in terms of proactive inhibition does not seem convincing. The forgetting would seem to occur during the counting activity, and since this should not be interfering on the basis of similarity, the finding is evidence for decay.

Studies of rate of presentation of sequences of digits for immediate recall have been considered evidence bearing on decay theory (Posner, 1963). A fast rate should result in less time elapsing from presentation to test of a given element of a sequence, and so the finding of better recall with a fast as opposed to a slow rate should be evidence for decay. However, other factors complicate the situation, notably the possibility for rehearsal which is greater with slow rates. Some studies have introduced another factor by permitting subjects to determine the order in which they produce written recall, as has been noted by Posner (1964). To the extent that recall has been found to be better with a fast rate of presentation and recall, in spite of rehearsal working in the opposite direction, the findings support a decay interpretation (Conrad & Hille, 1958).

In the case of paired associates Murdock (1963) has compared conditions in which pairs were presented for 1, 2, or 3 seconds, while time for recall of the cued response was held constant. No appreciable difference in the early part of the retention curve was found. On the other hand, Peterson (1963) found that a 2-second rate of presentation and test produced significantly better recall than a 4-second rate when one intervening event separated presentation from test. Thus, the early portion of retention curves from paired-associate experiments provides equivocal results in regard to the effect of rate. Of course, they agree in that a slow rate is not found to be better than a fast rate in the steepest part of the curve.

In summary, there is some evidence for the presence of decay, so that even though interference cannot be excluded as a source of short-term forgetting, there remains the possibility that deterioration with time is also involved. Whatever responses are available to the subject may compete at recall, if there is no basis for discrimination. These competing responses must be those which have not been forgotten since their presentation. The possibility exists that reduced availability of competing responses may occur in time whatever the intervening activity may be. An interpretation of this kind has been offered in connection with recall of individual trigrams (Peterson & Gentile, 1965). Retention was measured after varying periods of counting. Retention was found to be better the longer the interval from the end of one test to the beginning of the next. A reasonable explanation is that competition from the previous items decreased the longer the rest interval. There was evidence of fewer intrusions of a single letter from the previous trigram the longer the rest interval. However, this forgetting of the competing element did not itself seem to be due to interference. There is a suggestion that competition and deterioration with time may have cooperated in producing the effect.

The present conception of interference can be distinguished from a *displacement* interpretation of interference. It has been suggested (Bower, 1964; Waugh & Norman, 1965) that items presented earlier are displaced by later items. Such an interpretation may be adequate for curve fitting in various situations. However, evidence has been presented that time relationships are of some importance. It would seem further that the finding of spontaneous recovery would be difficult to account for in a displacement theory.

Duration of presentations appears to be of greatest importance in the later portion of short-term retention curves, where there is a positive relationship between duration of presentations and recall. Murdock (1963) has found that when from three to five pairs intervened between presentation and recall, recall was directly related to duration of presentation. Any effect of duration of the retention interval would seem to work against the direction of this finding, since the intervening pairs received longer presentations if the pair to be tested received such. In Peterson's (1963) rate experiment previously mentioned there was better recall for the slower rate when from two to six events intervened between presentation and recall. However, this superiority cannot unambiguously be attributed to the duration of presentation, since the latter was confounded with time permitted for recall.

The effect of longer durations of presentations in producing better performance at the asymptote of the short-term retention curve is in accord with studies of list learning in which longer presentation durations have also produced better performance (Bugelski, 1962; Nodine, 1963). The failure to find an advantage for longer presentation durations early in the short-term retention curve, when the most rapid forgetting occurs, suggests that something other than learning may be masking learning during this period. A recency mechanism is such a possibility.

REPETITIONS

Repetitions of presentations involve amounts of time devoted to the event to be remembered that are multiples of a single presentation. If increasing the duration of presentation improves recall, it would be expected that multiplying the number of presentations should also improve recall. This has been found to be the case, as is shown in Figure 3. The data plotted there are from a study of pairs of words and digits presented and tested at a 2-second duration per exposure (Peterson, 1963). Two pairings separated by an 8-second spacing interval produced better recall than one pairing after an 8-second retention interval. Intervals were filled with presentations and tests of other pairs. The failure of repetition to improve retention with a single intervening event and the improvement at the longer retention interval agree with findings by Murdock (1963). He tested lists of six pairs presented in the same serial order from one to three times. A strong effect of repetitions was found at the asymptotes of the retention curves, while there were only small differences in the initial slopes of the curves.

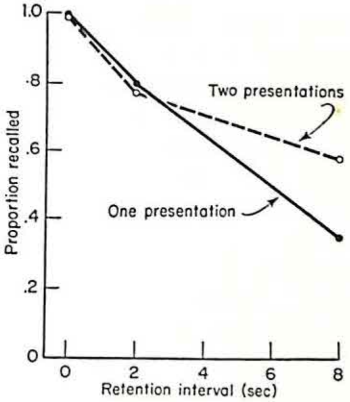

FIG. 3. Recall after one or two presentations of a pair.

When the effect of repetitions is considered, another variable must be taken into account. The time interval between repetitions has been found to be important in several studies (Greeno, 1964; Peterson, Hillner, & Saltzman, 1962; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963). In Figure 4 data

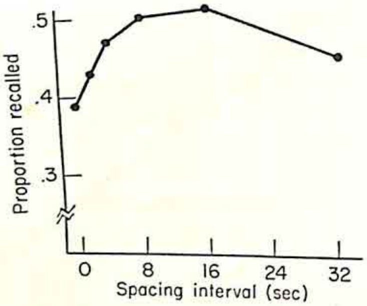

FIG. 4. Recall after a 16-second retention interval as a function of the spacing between two presentations.

have been plotted from a study varying the spacing interval, while retention was tested 16 seconds after the second presentation (Peterson, Wampler, Kirkpatrick, & Saltzman, 1963). Presentation and test of other pairs filled the intervals. Retention was found to improve as the spacing interval lengthened from 0–16 seconds. The decrease from 16–32 seconds was not significant, although it suggested that the beneficial effect of spacing has a limit.

The positive relationship between spacing and recall found at longer retention intervals contrasts with a slightly negative relationship at very short retention intervals (Peterson, Hillner, & Saltzman, 1962). Again, this finding of a differential effect of the same variable in different portions of the retention curve suggests that there is more than one mechanism in short-term retention.

The increasing spacing function at longer retention intervals contrasts with the decreasing retention function found over intervals of the same duration following a single presentation. A paradox appears in that the best way to learn a pair seems to be to allow the earlier pairing to be partially forgotten. One way to resolve the paradox is to assume that there are two components involved in storage over these intervals, and that with the passage of time the one diminishes while the other increases. An activity trace may diminish over the interval, while a structural trace is consolidated. At time of testing after two pairings, a longer time during which consolidation can occur has elapsed since the first pairing in the case of spaced pairings than in the case of massed pairings. The variable of duration is thus expanded to include time between pairings as well as time during pairings.

It was noted previously that experiments by Hebb (1961) and Melton (1963) indicate that a structural trace can be produced in a single trial, when interfering activity prevents any ongoing fixation process. However, this does not rule out the possibility that permitting a consolidation process to continue uninterrupted would result in a stronger structural trace. It would be assumed that the neural activity during a presentation persists in attenuated form when the presentation has ended, and in the present experiments that this activity can continue at the same time that the subject is engaged in other verbal activity not involving the same neural systems. It could not be expected that spacing would improve learning when activity during the spacing interval consisted of new arrangements of the same verbal elements used in the learning task. Such an arrangement was designed to produce maximum interference, and, indeed, Melton (1963) found the more other immediate memory tests that were given between repetitions of a given sequence of digits, the slower the learning was. Digits comprised the interpolated task as well as the learning task. In studies where improvement has been noted with spacing of repetitions the interpolated activity has involved elements differing from those of the learning task. In such a case the consolidation would not be pitted against massive interference.

A fixation process following learning has been inferred from a number of animal studies in which traumatic interruption of the hypothesized process has been investigated, and from drug studies in which facilitation or inhibition of the process has been reported (Gerard, 1963). In addition, Walker and his co-workers have incorporated the measurement of galvanic skin responses into studies of verbal retention with results interpreted as evidence for consolidation (Walker & Tarte, 1963). These studies have involved intervals of much longer duration than the intervals used to space presentations in the present studies. This does not preclude the possibility that the same process may be involved, but it suggests caution.

There has been a report of an electroconvulsive shock study with a substantial number of humans over intervals not differing greatly from those of the spacing studies previously described. Cronholm and Lagergren (1959) reported testing retention of a single presentation of a number when intervals of 5, 15, or 60 seconds elapsed from the end of presentation to the onset of shock. Later recall of the number was found to be better the longer the interval from presentation to shock. Nothing was said about the patient's verbal activity during this interval, and presumably he was free to rehearse. It would be of interest to have the results of a study in which controls for rehearsal were present, since the present application of the concept of consolidation requires that it occur at the same time that the subject is engaged in other verbal activity.

There are alternative explanations of the effect of spacing presentations. Greeno (1964) has suggested that spacing aids discrimination of a paired-associate item from other items by introducing these other items between repetitions. When presentations are massed, a second presentation occurs when generalization tendencies to produce errors are at a minimum, and so reduction of the generalization tendency will be small. While such an explanation might be relevant when cues of high similarity are used, it does not seem convincing as an explanation of studies in which words were used of low similarity with which the subject was very familiar.

Another possible alternative is the extinction-recovery theory which has been advanced to explain distributed practice effects in the learning and retention of lists (Keppel, 1964). It is held that associations (contextual and otherwise) are extinguished during presentations of later pairs, but that they recover with time after extinction to interfere with recall. Permanent extinction of competing associations is facilitated by distributed practice. Whatever the success of the theory in other contexts, it would not seem relevant for the present spacing experiments. Here the spacing and retention intervals were filled with the same kind of pairs. If these intervening pairs are assumed to produce extinction, then they cannot at the same time be assumed to permit recovery.

THE RECENCY MECHANISM

Further evidence for the notion of a short-term store as distinct from the learning mechanism is provided by the negative action of a recency mechanism during anticipation (Peterson, Brewer, & Bertucco, 1963). It was found in a paired-associate learning task that subjects in proceeding through a given trial tended to avoid using responses which had already been presented on that trial. Presumably the subject refrained from using as a guess a response which he knew must be wrong because of its recent presentation. The

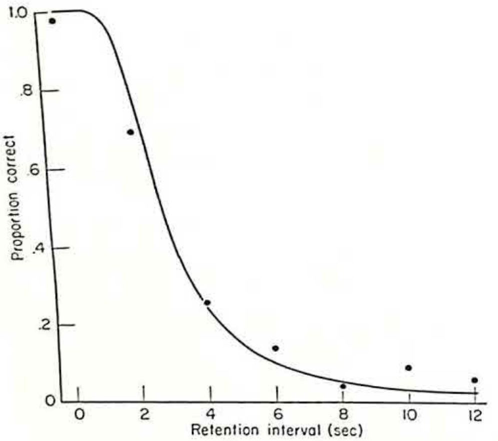

FIG. 5. Running memory span for pairs. (The curve is based on the equation $S(t) = 1 - \exp\left[-\frac{4(t+1)}{.12t^2(t+2)^2}\right]$.)

basis for the subject's recognition of an error would seem to lie in a recency mechanism. The action of the short-term mechanism during paired-associate learning may be characterized as a kind of running memory span. It is known that in a situation in which information enters the organism sequentially, the subject can report accurately, in order, the most recent digits or words that have been presented (Pollack, Johnson, & Knaff, 1959; Postman, Turnage, & Silverstein, 1964). It would appear that the subject has a similar possibility for recall in the continuous paired-associate situation as well. A test shortly after presentation should permit accurate recall on the basis of the running memory span without an association having been established. The results of an experiment[3] designed to investigate this possibility are summarized in Figure 5. Each of 20 students read aloud a sequence of pairs consisting of a word and a number presented at a 2-second rate in a memory drum. No tests were included among the sequences of pairs. The drum stopped after from 8–12 pairs had been presented, and a subject was given 10 seconds to speak all of the pairs that he could remember. Ten such lists of uncertain length were presented and tested with each subject. The recalls were scored in terms of the number of exposures of other pairs separating presentation of an individual pair and the beginning of the recall period. The ability of the subject to correctly recall a pair is seen to be very high immediately after presentation and to drop off quickly thereafter. It seems reasonable to assume that the mechanism that the subject uses under conditions of free recall can also be used when a cued recall is being tested. The sequence of previous presentations is immediately available to the subject for a few seconds in a short-term storage.

Note that the short-term store discussed here is a postperceptual store. Subjects read items aloud, and the retention interval is measured from that point. There is a distinction to be made between this short-term mechanism and that conceptualized by Sperling (1963) as the visual information storage. In Sperling's conceptualization the subject's rehearsal of information places it in an auditory information storage. The rapid decay of

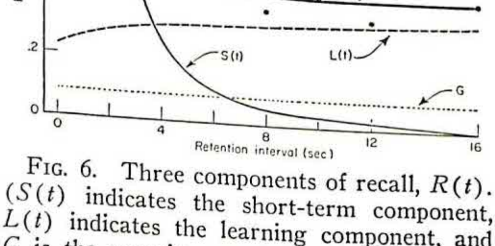

FIG. 6. Three components of recall, $R(t)$. ($S(t)$ indicates the short-term component, $L(t)$ indicates the learning component, and G is the guessing component.)

---

[3] Lawrence Clayton assisted in this experiment.

information that is found in his experiments using tachistoscopic presentation is assigned by him to the visual information storage.

INTERACTION OF COMPONENTS

Three components have been discussed in connection with the short-term recall of paired associates. A model of their interaction follows. The empirical probability of a correct recall, $R(t)$, is considered to be the union of the probabilities associated with the three components.

$$R(t) = S(t) + L(t)\,[1 - S(t)] + G\,[1 - S(t)]\,[1 - L(t)],$$

where $S(t)$ is the probability of a correct recall on the basis of the short-term mechanism, $L(t)$ is the probability of recall by action of the learning mechanism, and G is the probability of guessing correctly. The equation gives the probability that a correct recall will occur through the operation of at least one of the three components.

An application of the model to some data has been made in Figure 6 to illustrate the interaction of the components. Pairs consisting of a word and a number from 1–12 were presented in a memory drum and tested after an interval filled with presentations and tests of other pairs.[4] A 2-second rate was used. Primacy effects were controlled by the use of dummies in initial positions of sequences. Position in the sequence was balanced out over a number of sequences. The empirical points are the means of 12 subjects, each tested 32 times at each retention interval. The curve $R(t)$ has been fitted to the data points by a trial-and-error process. The contribution of the short-term mechanism, $S(t)$, is represented by the same curve used with free recall of pairs in Figure 5. The contribution of the learning mechanism, $L(t)$, has been made similar in form to a curve which would fit the data points for different lengths of spacing intervals in Figure 4. The guessing probability has been taken to be the reciprocal of the number of response alternatives.

The equations for $S(t)$ and $L(t)$ were chosen so that their combination with G would yield a curve giving a reasonable fit to the data points. The curves for the three components are, of course, hypothetical and not determined by the data of this experiment. Their forms have been inferred from the other experiments shown in Figures 4 and 5.

As Figure 6 illustrates, an increasing contribution of the learning mechanism can be hidden by a rapidly dropping short-term component. The learning component was presumed to have increased to some positive value by the end of the presentation, and thus it is greater than zero at the beginning of the retention interval. The point of view underlying the conceptualization is that recall is a search process which requires time. The strength or availability of a response due to the influence of one of the mechanisms is reflected by a latency distribution. The parameters of the distribution vary with time. Those of the short-term component result in a distribution whose mean becomes greater with time. In the case of the learning component the mean of the distribution decreases with time. The length of the recall period, 2 seconds in the present experiment, sets a limit on the maximum latency which can be successful. A condition whose latency distribution for the short-term component lies entirely within the recall period will have 100% correct recall. A retention condition whose range of latencies extends beyond the recall period will show forgetting. It follows that less would be forgotten if a longer recall period were provided, and hence it is not appropriate to speak of a pairing having been forgotten, or lost once and for all from storage. By this reasoning there is avoided the implausibility of a consolidation process resulting in something which had been lost later being produced out of nothing. Rather, it is possible to conceive of a trace which was present from the beginning being strengthened or becoming more readily available in the sense of requiring less time to be found.

---

[4] Richard Wampler assisted in this experiment.

It is not claimed that with unlimited recall time the subject would eventually hit upon the correct response. The recall period may itself be a source of interference. Further, the possibility of the subject stopping the search with an incorrect recall may increase with time. It is considered that with a short recall period this problem may for present purposes be ignored.

SUMMARY

The position taken in this paper acknowledges that short- and long-term retention have much in common. Interference produces forgetting in both cases, and repetition generally improves retention. However, it is considered that in the case of paired associates for a brief period after a presentation a recency mechanism interacts with the learning mechanism. Several variables affect the retention curve differently during the early declining phases as contrasted with the asymptote. Among these are duration of presentations, number of repetitions, and duration of the spacing interval between repetitions. The short-term or recency mechanism is conceptualized as a postperceptual mechanism whose effectiveness decreases through time and the action of other information. Recall for very recent events is characterized by rapid availability, by which is meant search times which can be effective in short recall periods. The long-term or learning mechanism is conceived of as a relatively stable store. There are indications that an increase in the availability of information on the basis of this component occurs for a period of time after presentation. In empirical retention curves the increasing component may be hidden by the rapidly decreasing short-term component. The presence of the former has been inferred from the beneficial effect of increasing the time between repetitions of a pairing and from the reminiscence found after irrelevant intervening activity.

REFERENCES

BOWER, G. H. An association model for response and training variables in paired-associate learning. *Psychological Review*, 1962, **69**, 34–53.

BOWER, G. H. Notes on a descriptive theory of memory. Paper read at New York Academy of Sciences Second Conference on Learning, Remembering and Forgetting, Princeton, September 1964.

BROADBENT, D. E. Flow of information within the organism. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 34–39.

BROWN, J. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 12–21.

BUGELSKI, B. R. Presentation time, total time, and mediation in paired-associate learning. *Journal of Experimental Psychology*, 1962, **63**, 409–412.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

CONRAD, R., & HILLE, B. A. The decay theory of immediate memory and paced recall. *Canadian Journal of Psychology*, 1958, **12**, 1–6.

CRONHOLM, B., & LAGERGREN, A. Memory disturbances after electroconvulsive therapy: III. An experimental study of retrograde amnesia after electroconvulsive shock. *Acta Psychiatrica et Neurologica Scandinavica*, 1959, **34**, 283–310.

ESTES, W. K. Growth and function of mathematical models for learning. In, *Current trends in psychological theory*. Pittsburgh: University of Pittsburgh Press, 1961. Pp. 134–151.

Friedman, M., & Gelfand, H. Transfer effects in discrimination learning. *Journal of Mathematical Psychology,* 1964, **1,** 204–214.

Gerard, R. W. The material basis of memory. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 22–33.

Goggin, J. Proactive and retroactive inhibition in the short-term retention of paired associates. Unpublished doctoral dissertation, University of California, Berkeley, 1964.

Greeno, J. G. Paired-associate learning with massed and distributed repetitions of items. *Journal of Experimental Psychology,* 1964, **67,** 286–295.

Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.

Hebb, D. O. Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning.* London & New York: Oxford University Press, 1961. Pp. 37–46.

Keppel, G. Facilitation in short- and long-term retention of paired-associates following distributed practice in learning. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 91–111.

Keppel, G., & Underwood, B. J. Short-term retention of paired-associate lists. Paper read at Midwestern Psychological Association, Chicago, May 1961.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior,* 1962, **1,** 153–161.

Loess, H. Proactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3,** 362–368.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 1–21.

Murdock, B. B., Jr. Short-term memory and paired-associate learning. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 320–328.

Nodine, C. F. Stimulus durations and stimulus characteristics in paired-associates learning. *Journal of Experimental Psychology,* 1963, **66,** 100–106.

Peterson, L. R. Associative memory over brief intervals of time. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 102–106.

Peterson, L. R. Reminiscence in short-term retention. *Journal of Experimental Psychology,* 1966, **71,** 115–118.

Peterson, L. R., Brewer, C. L., & Bertucco, R. A guessing strategy with the anticipation technique. *Journal of Experimental Psychology,* 1963, **65,** 258–264.

Peterson, L. R., & Gentile, A. Proactive interference as a function of time between tests. *Journal of Experimental Psychology,* 1965, **70,** 473–478.

Peterson, L. R., Hillner, K., & Saltzman, D. Time between pairings and short-term retention. *Journal of Experimental Psychology,* 1962, **64,** 550–551.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology,* 1959, **58,** 193–198.

Peterson, L. R., & Peterson, M. J. Minimal paired-associate learning. *Journal of Experimental Psychology,* 1962, **63,** 521–527.

Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. Recency and frequency in paired-associate learning. *Journal of Experimental Psychology,* 1962, **63,** 396–403.

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. Effect of spacing presentations on retention of a paired-associate over short intervals. *Journal of Experimental Psychology,* 1963, **66,** 206–209.

Pollack, I., Johnson, L. B., & Knaff, P. R. Running memory span. *Journal of Experimental Psychology,* 1959, **57,** 137–146.

Posner, M. I. Immediate memory in sequential tasks. *Psychological Bulletin,* 1963, **60,** 333–349.

Posner, M. I. Rate of presentation and order of recall in immediate memory. *British Journal of Psychology,* 1964, **55,** 303–306.

Postman, L., Turnage, T. W., & Silverstein, A. The running memory span for words. *Quarterly Journal of Experimental Psychology,* 1964, **16,** 81–89.

Sperling, G. A model for visual memory tasks. *Human Factors,* 1963, **5,** 19–31.

Walker, E. L., & Tarte, R. D. Memory storage as a function of arousal and time with homogeneous and heterogeneous lists. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 113–119.

Waugh, N. C., & Norman, D. A. Primary memory. *Psychological Review,* 1965, **72,** 89–104.

Wickens, D. D., Born, D. G., & Allen, C. K. Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2,** 440–445.

(Received December 30, 1964)

Psychological Review
1966, Vol. 73, No. 3, 208–231

REFERENTIAL PROCESSES OF SPEAKERS AND LISTENERS[1]

SEYMOUR ROSENBERG
Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey

AND

BERTRAM D. COHEN
Rutgers University

A stochastic theory is presented to (a) account for a speaker's selection of a linguistic response to distinguish a referent stimulus from nonreferent stimuli, and (b) predict the probability that a listener, using the speaker's response, correctly identifies the referent. The speaker's response is postulated to result from the concatenation of 2 hypothetical psychological stages, termed "sampling" and "comparison." The listener's identification of the referent is postulated to result from a 1-stage process similar to that of the speaker's comparison stage. Results from several interrelated experiments are reported which provide support for the basic assumptions of the theory. Psychological processes inherent in classical word association and in recognition and recall are interpreted in terms of the speaker and listener theory.

Philosophers, linguists, and psychologists generally find it necessary to introduce the concept of the referent, or closely related concepts, in their discussions of the semantic properties of language (e.g., Brown, 1958; Kantor, 1936; Morris, 1946; Ogden & Richards, 1923; Quine, 1960; Russell, 1921; Skinner, 1957). The concept of the referent includes the objects, properties, events, or relationships about which a person talks, writes, or gestures. Referents may be particular stimuli or classes of stimuli. Various writers include in the class of referents not only stimuli that are present in the immediate environment of a speaker or listener, but also objects and events that are removed in time or place, and even totally fictional constructions. In the present paper we restrict our inquiry to referents that are identifiable stimuli in the immediate environment of a speaker and a listener. With this restriction the psychological analysis of relations among the speaker's referent, the speaker's linguistic act, and the response evoked in the listener by the linguistic act becomes an empirically tractable problem.

The proposition, held by philosophers for centuries, that words do or should stand in a fixed relationship to their referents, is almost useless in the psychological analysis of human communication. Most referents have many names, and many referents can be called by the same name. A speaker's specific utterance is conditioned not only by his immediate social, physical, and linguistic context, but also by his linguistic development, previous interactions with the listener, perceptions of the listener, and by the particular response he intends to evoke from his listener. It is our contention that even when the effects of these variables are identified or controlled, there remains an important degree of unpredictability in the speaker's choice of a specific utterance; that is, the utterance of the speaker is a probabilistic choice from among a set of responses in the speaker's repertoire, each of which is more-or-less adequate as a linguistic cue for his referent. The process by which a particular utterance is selected and emitted is termed the referential process of the speaker.

---

[1] Supported in part by NSF Grant GS-179 to Bertram D. Cohen. We wish to acknowledge the valuable suggestions of Saul Sternberg who read an earlier draft of this paper. We also wish to thank Ernst Rothkopf for suggesting a number of the methodological refinements we used in collecting word-association data.

The interpretation of the speaker's utterance by a listener at any given time—or, more specifically, his identification of the speaker's referent—is also assumed to result from a probabilistic process. We term this the referential process of the listener. This is not to say that the listener's immediate environment, linguistic development, previous interactions with the speaker, perceptions of the speaker, and goals do not also play an important role in determining his reaction to the specific utterance of the speaker.[2]

One of the most serious handicaps in the investigation of referential processes in natural human communication is that the speaker's referent, even when it is in the immediate environment, is usually identifiable only by inference from the speaker's response. This method of identifying the referent makes any analysis of the process which links the referent to the speaker's response a circular one. Furthermore, the identification of the referent from the speaker's response may be unreliable if our conjecture about the stochastic nature of the process is correct.

The difficulties in identifying the speaker's referent also preclude a thorough analysis of the accuracy of the listener's decision process in referent identification. Added to this difficulty is the fact that overt discriminative listener reactions to the speaker's referent often do not (or cannot) occur at all in conversation or in reading. It is not surprising, therefore, that the investigation of referential processes by psycholinguists has been primarily anecdotal—a few examples of referents and linguistic responses typically bear the burden of relating theory to linguistic events.

The purposes of the present paper are to (*a*) describe an experimental paradigm in which the speaker's referent is identifiable independently of the speaker's behavior, (*b*) describe the main features of a stochastic theory of referential behavior, and (*c*) summarize a series of experiments designed to test basic propositions of the theory. A number of implications of the theory for certain well-establised verbal phenomena will also be sketched.

EXPERIMENTAL PARADIGM

The basic features of the experimental paradigm are as follows. A referent stimulus is assigned to the speaker by the experimenter. The speaker's task is to provide a linguistic response that makes it possible for his listener to identify the referent stimulus. Given the speaker's response, the listener's task is to identify the referent.

An important additional feature of the situation is the fact that the nonreferents from which the referent is to be distinguished are made explicit to the speaker. Also the listener is given an explicit set of stimuli from which to choose the speaker's referent. The use of well-defined stimulus sets for both speaker and listener is a convenient simplification of the task and not a necessary feature of the experimental investigation of referential behavior.

In the studies described in this paper, the experimental paradigm was realized in the following manner. The speaker and the listener were both shown pairs of words, one word pair at a time. The experimenter designated one member of each word pair as the speaker's referent and the other as the nonreferent. Note that in these initial studies the referents and nonreferents were themselves words.[3] The speaker was instructed to provide a one-word clue (the *speaker response*) that would enable the listener to identify the referent word correctly. The speaker was told to use a "clue word that has some understandable connection with" the referent and *not* to use any "clue word that sounds like or is spelled like" either of the stimulus words. The listener was instructed to identify the referent word from the two alternative possibilities using the speaker response.

---

[2] "Speaker" will be used in the generic sense in this paper to refer to a person whose communicative act is oral, written, or gestural. "Listener" will be used in a similar generic sense.

Consider, for example, the synonym pair WOMAN-LADY with LADY as the referent. Examples of permissible and effective (for listener's referent identification) speaker responses include *tramp, title, Chatterly, finger,* and *pink.* A permissible but probably ineffective speaker response is *female.* Most of the word pairs used in the present series of studies were synonym pairs. The choice of synonym pairs was motivated by the thought that the responses associated with the referent are likely to overlap considerably with those of the nonreferent and that the effects of the nonreferent on speaker behavior are most salient under conditions of high overlap. A number of unrelated word pairs, for example, BUY-ATTACK, were also included. The method is, of course, amenable to almost any type of word pair, for example, supraordinates (ANIMAL-DOG), coordinates (DOG-CAT), pairs related in experience (TABLE-CHAIR), etc.

The experimental method is also readily adapted to nonverbal stimuli, more than one nonreferent stimulus, and complex speaker responses (phrases, sentences, etc.). A number of investigators (Carroll, see Osgood & Sebeok, 1954, p. 200; Krauss & Weinheimer, 1964; Maclay, 1962; Maclay & Newman, 1960; Werner & Kaplan, 1963) have devised and used experimental situations for studying the communication of nonverbal stimulus objects (blocks and pegs varying in shape, size, and color; simple geometric forms; novel designs; olfactory stimuli). The speaker's task is essentially to describe one of the objects so that the listener can select it from among a set of alternatives. Lantz and Stefflre (1964) recently proposed that the accuracy with which a color is communicated to listeners be used as a measure of the codability of the color. This measure was found to be the best of several measures of stimulus codability in terms of its ability to predict the ease with which the colors could be remembered in recognition tasks. To our knowledge, the model we propose is the first systematization of the psychological processes governing speaker and listener behavior in this type of communication task.

THEORY

Consider a population of subjects in which differences among them in their subcultural backgrounds have been minimized. That is, all speakers and listeners have been drawn from a population of persons who are similar in age, education, and language abilities. It is also convenient, at this point, to restrict our attention to speaker-listener pairs who, though similar to one another, are strangers. This restriction is imposed in order to minimize the special effects on communicative behavior that various types of interpersonal relationships might produce. How shall we account for the variability among these speakers in the responses they emit to communicate a referent? Why do some, but not all, of the listeners correctly identify the referent given a particular speaker response?

One approach is to account for the behavioral variability of speakers and listeners in terms of the similarity in meaning of the referent and nonreferent words—or, in the case of nonverbal stimuli, by similarities in their names or in their physical properties. Thus, one might expect that both the number of different speaker responses given by a group of subjects and the probability of a listener error are positively related to the similarity in meaning of the referent and nonreferent. Measures of meaning similarity can be derived from semantic-differential measures or from the overlap of word associates (or names) given to the referent and to the nonreferent stimulus word (Marshall & Cofer, 1963). How can these measures be used to predict details of speaker and listener behavior, for example, the specific set of speaker responses and their probabilities, and the probability of a listener error for any particular speaker response? A conceivable approach (when the stimulus array consists of words) is to obtain an overlap measure between each of the stimulus words and each of the possible speaker responses. Speaker responses could then be ordered in terms of their relative overlap with referent and nonreferent; from this ordering one might be able to estimate the distribution of speaker responses and, for any given speaker response, predict the probability of a listener error. Unfortunately, this approach is unwieldy since it requires the collection of word-association data for *each* of the speaker responses and for each of the stimulus words in the communication task. It is also somewhat circular in that it requires that we identify the speaker responses before we can estimate their probabilities.

---

[3] Writers concerned with referential behavior often distinguish between words (signs) and "things" (referents). From a psychological standpoint, it is important to maintain a distinction between speaker responses and referents. Any stimulus, however, can function as a referent. There is no reason to restrict the term referent to nonverbal stimuli. Thus, spoken or printed words, as well as nonverbal stimuli, can and often do function as referents. The advantage of using verbal rather than nonverbal referents in initial studies will be made apparent later in the paper.

An alternative approach is considered here. Results from preliminary work suggested that standard word-association norms for the referent and nonreferent stimuli alone are sufficient to predict details of speaker and listener behavior using the stochastic theory outlined below. The essential notions of the theory are as follows.

It is postulated that the speaker's referential process is a concatenation of two hypothetical psychological stages, termed *sampling* and *comparison*. Both stages are assumed to be probabilistic. The speaker's referential process begins with a sampling stage in which a response from the speaker's repertoire is selected randomly from the distribution of word associations to the referent stimulus alone. The distribution of associations to the referent is denoted I_r. In the comparison stage, which always follows sampling, the relative "associative strengths" of each of the stimulus words to the sampled response determines the conditional probability that the speaker will utter the sampled response. The associative strengths of the referent and nonreferent to a sample response, i, are denoted $s_i(r)$ and $s_i(n)$, respectively. If, as a result of the comparison stage, the sampled response is rejected as an inadequate communication, the sampling-comparison cycle is repeated. Ultimately the speaker emits a response, thus terminating the process. A diagram of this hypothetical process is given in Figure 1.

It is assumed that the listener's choice is the result of a probabilistic process which is a function of the associative strengths of the speaker response to each stimulus word. For the listener the associative strengths of the referent and nonreferent to a speaker response i are denoted $l_i(r)$ and $l_i(n)$, respectively. The probability that the listener will correctly identify the referent is determined by these two associative strengths (see Figure 1).

The referential process of the listener is assumed to be similar to the comparison stage of the speaker. Put another way, the speaker essentially takes the role of the listener in the speaker's comparison stage—a notion reminiscent of Mead's (1934, p. 147) use of the "generalized other" in discussing language.

The formal theory is not necessarily limited to verbal stimuli. The choice

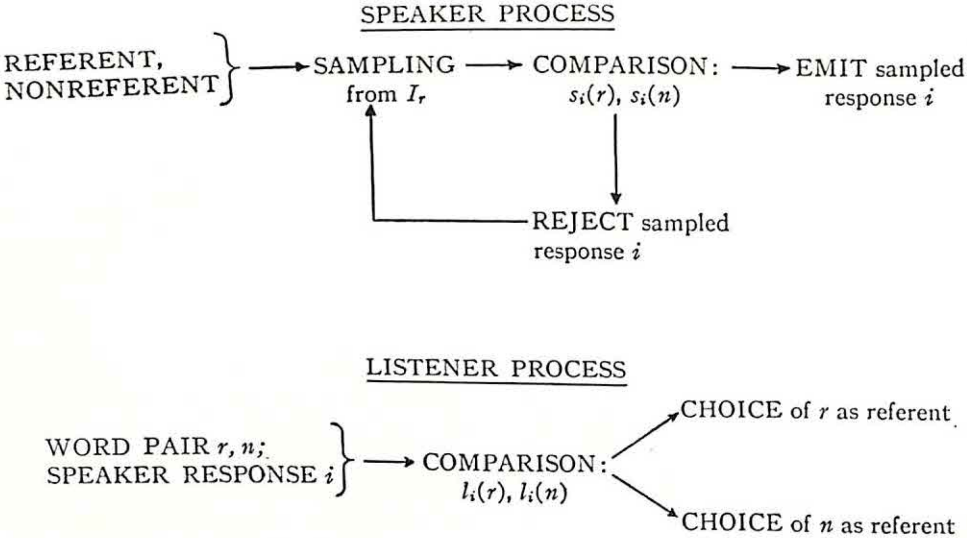
SPEAKER PROCESS
REFERENT, NONREFERENT
SAMPLING from I_r
COMPARISON: $s_i(r)$, $s_i(n)$
EMIT sampled response i
REJECT sampled response i
LISTENER PROCESS
WORD PAIR r, n; SPEAKER RESPONSE i
COMPARISON: $l_i(r)$, $l_i(n)$
CHOICE of r as referent
CHOICE of n as referent

FIG. 1. Sketch of the speaker and listener processes.

of verbal, rather than nonverbal, stimuli is particularly advantageous for these initial studies if word-association data provide satisfactory estimates of the relevant quantities used in the theory. Nor is the theory necessarily limited to single-word speaker responses. Phrases, sentences, etc., may be permitted as speaker responses. However, the operations analogous to word-association techniques for nonverbal stimuli and for multiple-word responses are not yet well developed.

Speaker's Sampling Stage

A response belongs to I_r if there is a nonzero probability that it will occur as a word associate to r. The probability $P_r(i)$ of sampling response i in I_r is not uniform for all responses in I_r as it might be in sampling from a section of a thesaurus. Since the speaker must sample from I_r more than once if a sampled response is rejected in the comparison stage, the question arises concerning the probability of again sampling the rejected response. A plausible and mathematically tractable assumption is that sampling is done with replacement so that $P_{r,t}(i) = P_{r,t+1}(i) = P_r(i)$ for all t and i. The additional subscript t indicates the tth sampling which has followed $t-1$ rejections of i. Our subjective experience with the speaker's task suggests that in searching for an adequate response he may consider the same response more than once and may, in fact, emit it after having rejected it one or more times.[4]

Speaker's Comparison Stage

Each response i sampled by the speaker defines, for a given referent-nonreferent pair, a pair of scale values, $s_i(r)$ and $s_i(n)$, respectively. When there is a high degree of overlap between the associative responses to the referent and nonreferent, as is the case with synonyms, the difference between $s_i(r)$ and $s_i(n)$ will be relatively small for many of the sampled responses. The closer the values of $s_i(r)$ and $s_i(n)$, the more likely it is that the speaker will make an incorrect decision in the comparison stage. For example, a speaker might emit a sampled response i although the associative strength of stimulus r to response i is less than that of n to i.

---

[4] Another plausible scheme is that sampling is done with replacement, but $P_{r,t}(i) > P_{r,t+1}(i)$ if i was sampled and rejected at time t. A third alternative is that $P_{r,t+1}(i) = 0$ if i was sampled and rejected at time t. These two sampling schemes require an additional rule for readjusting the probabilities of sampling the other responses in I_r. In the third sampling scheme, a speaker may exhaust I_r unless I_r is infinite.

As we have already noted, a speaker might reject response i at one time and, having sampled i again at a later time, might then emit it. It is explicitly assumed that the conditional probability of emitting a sampled response is unchanged by previous instances of sampling and rejection.

The identification of the comparison stage as a decision process suggests a number of models that have been developed for choice behavior and comparative judgment. A prominent model, applicable to the speaker's comparison stage, is Thurstone's model of comparative judgment. According to this model, speaker comparison consists of selecting a value from each of two hypothetical distributions of associative strength, say y_r and y_n; if $y_r > y_n$, the speaker emits the sampled response; otherwise he rejects the response and samples again from I_r. A second model, the one we propose to use for the speaker's comparison stage, was first suggested by Bradley and Terry (1952). In spite of the formal differences between the Thurstone model (Case V) and the Bradley-Terry model, they are almost equivalent in practice (Torgerson, 1958, p. 202).

The essential notions of the Bradley-Terry model are simple and straightforward and require fewer assumptions in practice than the Thurstone model. Another desirable feature of the Bradley-Terry model is its link to Luce's (1959) theory of choice behavior. The usefulness of this link will become apparent when we extend our analysis of referential behavior to multiple nonreferents.

The application of the Bradley-Terry model requires that $s_i(r)$ and $s_i(n)$ be restricted to values equal to or greater than zero for all i, r, and n. Let $P_c(i|i)$ be the conditional probability of emitting response i, given that i is sampled. Then, for $P_c(i|i) \neq 0,1$,

$$P_c(i|i) = s_i(r)/[s_i(r) + s_i(n)]. \quad [1]$$

What remains to be done is to specify precisely how the associative strengths, $s_i(r)$ and $s_i(n)$, are obtained. A relationship is postulated between associative strength and associative probability obtained from word-association data. This relationship is given below in the section on Estimation of Theoretical Quantities.

Asymptotic Distribution of Speaker Responses

If the probabilities $P_r(i)$ and $P_c(i|i)$ remain unchanged with repetitions of the sampling and comparison stages (as assumed in the previous sections), then the entire speaker process is a stationary Markov chain. There is an observable absorbing state corresponding to the emission of each response in I_r and a single unobservable transient state, T, corresponding to the rejection of any response in the comparison stage.

Thus, the probability that a speaker emits any given response as a result of a single sampling-comparison cycle is the product of two probabilities: the probability that the response is sampled, and the conditional probability that it is emitted given that it is sampled. That is, the transition probability:

$$P_r\{E_{i,t+1} \text{ (emission of response } i \text{ at time } t+1) | T_t\} = P_r(i)P_c(i|i).$$

Now, if the speaker has not emitted a response, the probability of rejecting any response in I_r following a single sampling-comparison cycle is:

$$P_r\{T_{t+1}|T_t\} = 1 - \sum_i P_r(i)P_e(i|i).$$

The initial probability vector for the states $T, E_1, E_2, \ldots$ is $[1,0,0,\ldots]$.

If the speaker is given an unlimited time to emit a response, then it can be shown (Kemeny & Snell, 1960, pp. 43–68) that the asymptotic probability of emitting response i,

$$p_i = P_r(i)P_e(i|i)/\sum_{j\epsilon I_r} P_r(j)P_e(j|j). \quad [2]$$

Listener's Comparison Stage

The application of the Bradley-Terry model to the listener's identification of the referent is as follows. Let x and y denote the two stimulus words, and let $P_i\{x,y\}$ denote the probability that a listener identifies x rather than y as the referent given speaker response i. Then if $P_i\{x,y\} \neq 0,1$,

$$P_i\{x,y\} = l_i(x)/[l_i(x) + l_i(y)]. \quad [3]$$

Equation 3 has the same form as Equation 1. The associative strengths $l_i(x)$ and $l_i(y)$ are always restricted to values equal to or greater than zero.

We have, however, deliberately introduced different notation (l and s) to refer to the associative strengths in the listener's and in the speaker's comparison stages. Although the speaker's comparison stage is conceptualized as one in which he essentially takes the role of the listener, the associative strength between any given stimulus x and response i in the speaker's comparison stage may not be equal to that involved in the listener's comparison stage. It would, however, be expected that $l_i(x)$ and $s_i(x)$ are monotonically related. Estimation of $l_i(x)$ and $l_i(y)$ from associative probability is given in the next section. It may also be noted that $P_i\{x,y\}$ in Equation 3 can be estimated directly from listeners' data (providing they are homogeneous), whereas $P_e(i|i)$ in Equation 1 is a theoretical quantity which cannot be directly estimated from speakers' response data.

Multiple Nonreferents

A generalization of the theory to communication involving more than one nonreferent will now be sketched. Consider, first, the referential process of the speaker. We retain the assumption that the effects of the nonreferents on the choice of a speaker response take place in the comparison stage only. Thus the sampling stage remains unchanged for multiple nonreferents. A generalized version of the Bradley-Terry model (Luce, 1959, p. 23), applied to the speaker's comparison stage, is as follows. If $P_e(i|i) \neq 0,1$, then:

$$P_e(i|i) = s_i(r)/\sum_{x\epsilon X} s_i(x), \quad [4]$$

where X is the set of referent and nonreferent stimuli.

The extension of the listener's referential process to a choice from among more than two stimuli takes the same form as that for the speaker's comparison stage. That is, the probability that the listener will identify any $x\epsilon X$ as the referent is given by:

$$l_i(x)/\sum_{y\epsilon X} l_i(y).$$

However, since Luce's (1959) generalization of the Bradley-Terry model is a consequence of Luce's choice axiom, it is also of interest to consider the possibility that the choice axiom itself (sometimes called the constant-ratio rule) also holds for listener behavior in our experimental paradigm. For an empirical test of the choice axiom we considered the postulated relationship between two-choice and three-choice data. That is, the following form of the choice axiom was used: If $P_i\{x,y\} \neq 0,1$, then:

$$P_i\{x,y\} = \frac{P_i\{x,y,z\}}{P_i\{x,y,z\} + P_i\{y,x,z\}}, \quad [5]$$

where $P_i\{x,y,z\}$ and $P_i\{y,x,z\}$ are the probabilities of choosing x and y, respectively, from the stimulus array x,y,z. (The last two stimulus elements in brackets can be written in either order.) Equation 5 says, in effect, that the probability that a listener chooses x as the referent when only x and y are the alternative possibilities is the same as the conditional probability of choosing x from x and y when $x,y,$ and z are available; the speaker response is, of course, the same in both cases. It should be noted that this test of the model is parameter free and does not require word-association data.

Estimation of Theoretical Quantities

The models discussed in the preceding sections apply to the behavior of an individual. However, in order to test these models, it is necessary to use grouped data. The use of grouped data implies a number of intersubject homogeneity assumptions. In order to satisfy these homogeneity assumptions, the subjects selected for experimental tests of the theory were restricted to male undergraduates with relatively similar academic interests and within a narrow age range.

It is assumed that the probability that a listener chooses a stimulus as the referent, for a given stimulus array and speaker response, is the same for all listeners. Thus, the proportion of listeners who, given speaker response i, chose x from an x,y array or from an x,y,z array was used to estimate $P_i\{x,y\}$ and $Pi\{x,y,z\}$, respectively. It is also assumed that the proportion of subjects who give i as a word associate to stimulus x in a standard word-association situation is an adequate estimate of $P_x(i)$, the probability that i is sampled by a speaker during the sampling stage, given x as the referent.[5]

Word-association data also enter into the calculation of associative strengths for the speaker and the listener. In the case of the listener, it is assumed that for any stimulus x,

$$l_i(x) = a[P_x(i) + c]^b, \quad [6]$$

where a, b, and c are constants, and where $P_x(i)$ is estimated by the proportion of subjects who give response i to stimulus x in the word-association task. The same power function is also assumed to hold for the speaker's associative strength $s_i(x)$, but possibly with different constants, that is, $s_i(x) = A[P_x(i) + C]^B$.

The use of a power transformation is suggested by analogy of $P_x(i)$ with a physical continuum, that is, from the power law suggested by Stevens (1957) for psychophysical data. Whether or not the analogy is defensible, the appropriateness of the power transformation for the present data is an empirical matter. This and other transformations were examined and compared using listener data. The results of these comparisons are given in the section on Tests of the Listener Model.

Metareferential Behavior

The experimental paradigm and possibly the theory can also be extended to the study of speaker and listener responses that are typically found in human communication as auxiliary to or consequences of the referential function. For example, a listener may reject a speaker's referential response as inadequate ("I don't understand."), solicit additional speaker behavior ("Tell me more about it."), or confirm a tentative guess about the referent by supplementing the speaker's response ("Do you mean the *red* book?"). The relationship of these metareferential functions (or metalingual functions as they are termed by Jakobson, 1960, p. 375) to the referential function can be studied experimentally by suitable additions to the responses available to the listener. Evidence from a preliminary study suggests that the probability that a listener rejects a speaker response as inadequate for the identification of the referent is positively related to the similarity between the associative strengths of the speaker response to each of the two stimulus words.

---

[5] Note that it is not necessary to assume that the probability with which an individual subject gives i as a *word associate* to stimulus x is the same as the $P_x(i)$ for that subject. We shall comment further on this issue below in Intersubject Homogeneity in Word Association.

COLLECTION OF WORD-ASSOCIATION DATA

Word-association data were collected from 544 male undergraduates from two introductory psychology sections and one introductory music section at Rutgers University. Two hundred and seventy-two subjects were given one member from each of 153 synonym pairs as stimuli; the remaining 153 words were given to the 272 additional subjects. Subjects were tested in groups of 100–200 students each.

Each subject was given a deck of IBM cards consisting of instruction cards and 153 cards with a different stimulus word printed on each card. The stimulus words given to each subject were in a different random order.

Subjects were asked to print one and only one word on each card—"the first word that comes to mind" for each stimulus word.

Each response was punched on the same IBM card that the subject used. Spelling errors were corrected. Response-frequency counts for each of the 306 stimulus words were obtained directly from the punched cards with the 7090 computer.

TESTS OF THE LISTENER MODEL

The two-person situation, in which each listener is assembled with only one speaker, is an inefficient way to test the various assumptions and consequences of the listener model. From the standpoint of obtaining stable estimates of $P_i\{x,y\}$ or $P_i\{x,y,z\}$, the most desirable set of data is one in which a sizable number of listeners is given the *same* speaker response for any given stimulus array. Furthermore, when listener behavior is linked with word-association data, it is desirable to base parameter estimates (e.g., estimates of a, b, and c in Equation 6) on a large number of x,y,i triples which vary over a wide range of $P_x(i)$ and $P_y(i)$ values and combinations. In short, the fortuitous set of speaker responses provided by each speaker to his listener does not have these desirable properties.

Listener Behavior and Associative Strength

Instead of a two-person situation, 110 male undergraduates (not previously used) were all given the listener task with the same set of 256 x,y,i triples. Each triple consisted of a synonym pair and a speaker response. The subjects were run in one of four groups of 25–30 persons, and each group was given the 256 word triples in a different random order. Several synonym pairs appeared more than once in the list but with different speaker responses. Each word triple was projected on a screen for 8 seconds.

Subjects were told that speaker responses had been obtained from college students during a previous semester and that their task as listeners was to identify which of the two synonyms in each word pair had been the speaker's assigned referent. Listeners were actually given speaker responses systematically selected from among the word associates of one or both members of the synonym pairs in order to obtain a wide range of $P_x(i)$ and $P_y(i)$ values. The relevance to the two-person situation of listeners' data obtained from this study depends on the extent to which actual speaker responses to a referent are, in fact, the same as word associates to that referent—as the speaker model assumes. As we shall see in the studies concerned with speakers, speaker responses and word associates to a referent do overlap considerably.

A convenient form of the model for fitting it to data is as follows. If Equation 6 is substituted into Equation 3 and the appropriate algebraic manipulations are made, the following linear equation with zero intercept is obtained:

$$\log [P_i\{x,y\}/P_i\{y,x\}] = b \log \{[P_x(i)+c]/[P_y(i)+c]\}, \quad [7]$$

where $P_i\{y,x\} = 1 - P_i\{x,y\}$. The proportion of listeners who chose x was used as an estimate of $P_i\{x,y\}$ in Equation 7. Also, the proportions of times any speaker response i occurred as a word associate of x and of y were used as estimates of $P_x(i)$ and $P_y(i)$, respectively, in Equation 7.

An iterative least-squares procedure resulted in the following estimates: $\hat{b} = .700$, $\hat{c} = .0018$. These estimates[6] were based on 255 word triples with the notation x and y assigned to each synonym so that $P_x(i) \geq P_y(i)$—one triple was deleted because $\hat{P}_i\{x,y\} = 1.00$. Figure 2 is a plot of the data with the fitted line having a slope equal to $\hat{b}$. There is no indication of any systematic curvilinearity in the observed relationship. These findings, which are in support of the model, encouraged us to examine the listeners' data in further detail.

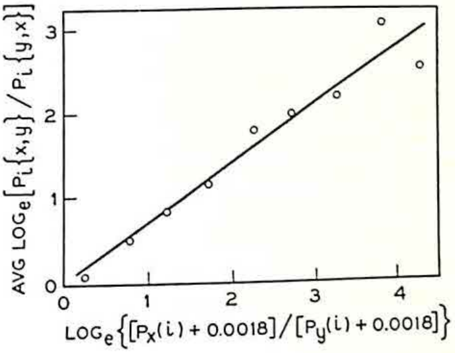

FIG. 2. The use of a single-valued function of $P_x(i)$ and $P_y(i)$ derived from the comparison model to relate word-association data and listeners' choice data.

It is important to ascertain, for example, how well the model, with parameters b and c estimated from all the data, fits selected subsets of the data. The generality of the model is considerably diminished if the constants in the power function depend on the values of $P_x(i)$ and $P_y(i)$. Therefore, the data were first partitioned on the basis of $P_x(i)$ and $P_y(i)$ into the following four subsets: (*a*) $P_x(i) \geq 0$, $P_y(i) = 0$; (*b*) $P_x(i) > 0$, $0 < P_y(i) \leq .012$; (*c*) $P_x(i) > .012$, $.012 < P_y(i) \leq .030$; (*d*) $P_x(i)$, $P_y(i) > .030$.

Data from each subset were then fitted separately to Equation 7 using the iterative least-squares procedure. A minimum sum of squared deviations was obtained for each subset. These minimum sums of squared deviations decreased by .59%, 2.84%, .56%, and .95%, respectively, from the four sums of squared deviations based on $\hat{b} = .700$ and $\hat{c} = .0018$. (An increment is not possible.) The average decrease is not statistically significant. Thus the parameters, b and c, do not appear to depend strongly on the combinations of $P_x(i)$ and $P_y(i)$ values.

Let us now consider two special cases (simplifications) of the model.

The first is the case where b is assumed to be 1; that is, a simple linear function is used in place of the general power function to estimate associative strength from associative probability. To test the adequacy of this special case, the data from all 255 word triples were fitted to Equation 7 by means of a least-squares procedure. The resulting minimum sum of squared deviations is 8.78% larger than the minimum obtained from the two-parameter $(\hat{b},\hat{c})$ model. This increase is significant beyond the .001 level. These results appear to rule out a simple linear relationship between associative probability and associative strength.

---

[6] Since a slight linear relationship appears to exist between values of the independent variable and the variance of these values, b was also estimated by the method of averages (Snedecor, 1946, pp. 105–108). The resulting $\hat{b}$ was .69.

The second simplification of the general model is to delete c from the power function, that is, set $c = 0$. This simplification seems particularly reasonable since $\hat{c}$ is so close to zero. Note, however, that when $P_x(i)$ or $P_y(i)$ equal zero, then $l_i(x)$ or $l_i(y)$ equal zero (see Equation 6) so that Equation 3 does not satisfy the restriction that $P_i\{x,y\} \neq 0,1$. If this restriction were removed from Equation 3, then the predicted $P_i\{x,y\}$ is 1 when $P_y(i) = 0$ for all values of $P_x(i) > 0$. However, in the data $P_i\{x,y\}$ is a monotonically increasing function of $P_x(i)$ when $P_y(i) = 0$ (Rosenberg & Cohen, 1964)—a result that is in accord with the model if $c > 0$.

Before accepting the use of the power function, one might wish to investigate a reasonable alternative. It may be argued, for example, that the probability that a listener chooses x as the referent in an x,y pair is a monotonic function F of the difference between the associative probabilities $P_x(i)$ and $P_y(i)$, that is,

$$P_i\{x,y\} = l_i(x)/[l_i(x) + l_i(y)] = F[P_x(i) - P_y(i)]. \quad [8]$$

It has been shown (Adams & Messick, 1957) that Equation 8 implies that an exponential function relates associative probability to associative strength, that is,

$$l_i(x) = h \exp [kP_x(i)], \quad [9]$$

where h and k are constants. Substituting Equation 9 into the Bradley-Terry model (Equation 3) and making the appropriate transformations yields the following one-parameter linear equation:

$$\log [P_i\{x,y\}/P_i\{y,x\}] = k[P_x(i) - P_y(i)]. \quad [10]$$

The sum of squared deviations from a linear least-squares fit using Equation 10 was almost double (92% greater) that of the sum of squared deviations involving the power function. Unfortunately, there is no satisfactory statistical test for comparing these two sums of squared deviations. Nevertheless, the sizable difference in squared deviations and the fact that it is based on a relatively large number of observations warrant a preference for the power transformation.

Test of the Choice Axiom[7]

A group of 236 male undergraduates (not previously used) were each given 48 sets of three stimuli along with a speaker response for each stimulus triple. The same subjects were also given the three possible stimulus pairs formed from a set of three stimulus words along with the same speaker response. Half the subjects completed the three-choice part first, and the other half completed the two-choice part first.

The following example from this study may help to clarify the empirical test of the choice axiom. Let x, y, and z be the stimulus words LIBERTY, LAWFUL, and FREEDOM, respectively; the speaker response i is *equality*. From the presentation of all three stimuli, the following proportions were obtained for choices of x, y, and z, respectively: .538, .103, .359. Using Equation 5 it would be predicted, for example, that when only x and y are given with the speaker response, $P_i\{x,y\} = .538/(.538 + .103) \cong .839$. When only x and y were presented with the speaker response *equality*, the obtained estimate of $P_i\{x,y\}$ was found to be .834. Similarly, predictions can be made from the three-choice data to the proportions of choices of y from y, z and of x from x, z.

Figure 3 summarizes the results obtained from the entire set of data. There appears to be no systematic deviation of the set of points from the diagonal line. A least-squares fit of the points yielded a regression coefficient of .94. The correlation coefficient between predicted and observed values was .97. In short, these data strongly

---

[7] Data for this section were collected by Austin Ely.

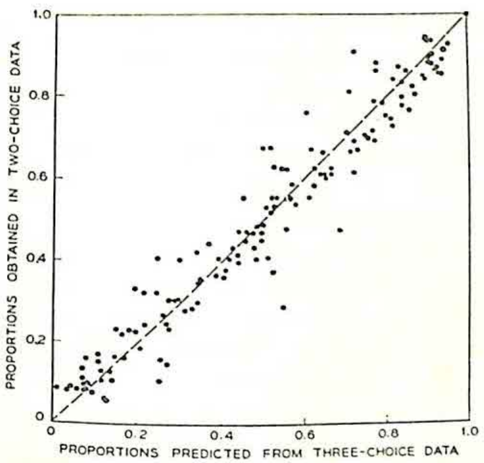

FIG. 3. Scatter diagrams of observed proportions of referent choices in two-choice arrays and the predicted proportions obtained by Luce's choice axiom from the observed proportions in three-choice arrays. (Each point represents the data from an individual set of three stimulus words and a speaker response.)

support the proposition that listeners' choices conform to the axiom.

As we have already noted, estimates of associative strength are not required in this test of the model. Thus, it is possible to extend this form of the listener model to stimulus materials consisting of complex nonverbal stimulus arrays and to speaker responses consisting of phrases, sentences, and nonverbal communications.

TESTS OF THE SPEAKER MODEL

This section describes a number of empirical tests of the speaker theory. Unless it is otherwise noted, these empirical analyses were based largely on data collected from 272 male undergraduates (not previously used). Each subject was given a deck of IBM cards consisting of instruction cards and 50 stimulus cards. A word pair with the referent underlined was printed on each stimulus card. Forty word pairs were synonyms, and 10 were unrelated word pairs. The subjects were all given the same 50 word pairs, but in different random orders. They were instructed to print a one-word speaker response on each card before going on to the next card. Subjects were told that their responses along with the word pairs would be shown to listeners at a later date and that their "scores" would depend on their listeners' accuracy in identifying the referents.

Results from two other speaker studies are also cited briefly in this section. These studies will be described when they are used.

Consequences of the following fundamental assumptions were investigated: (*a*) *Intersubject homogeneity*—Given any referent r, all subjects sample from the same response set I_r; (*b*) *Word associates and speaker responses*—Word associates of the referent stimulus provide an empirical identification of the responses in I_r; (*c*) *Comparison stage*—The comparison stage depends on the relative associative strengths, $s_i(r)$ and $s_i(n)$, of a sampled response i to each of the stimuli, r and n.

Empirical consequences derived from the first two assumptions are parameter free and are equally valid for the stationary and nonstationary models discussed in the Theory section. The validity of these first two assumptions obviously depends on the way subjects are selected.

Analyses concerned with assumption (*c*), that is, the nature of the comparison stage, depend on the assumption that word-association norms provide a satisfactory estimate of the probability that i is sampled by the speaker in the sampling stage. Generally speaking, the sampling and comparison stages are sufficiently intertwined so that it is difficult to study details of one stage without making explicit assumptions about the other.

Intersubject Homogeneity

The ideal way to determine the similarity among response sets I_r of a group of speakers is first to identify the responses in each speaker's I_r. One method of identifying these responses is to require a speaker to emit a number of responses to each referent, a method analogous to continued associations. A related method is to intersperse presentations of the same referent with time periods filled with other activities so as to reduce sequential dependences among the speaker responses to the referent. It is questionable whether either of these methods yields an adequate sample of the responses in an individual's I_r.

An alternative, albeit less direct, test of the homogeneity assumption is as follows. If all subjects sample randomly from the same I_r (and then go through the same comparison stage), then the preference for a specific kind of speaker response, for example, synonym, antonym, common responses, etc., is not expected to be a consistent property of an individual across items (as it often is in the word-association task). Of these various idiosyncratic preferences, the preference for common responses would appear to be one least disturbed or suppressed by the communication task. That is, an experimental test involving the speaker's use of common responses would appear to be the strongest test of the homogeneity assumption.

Two measures of communality were calculated for each speaker and for each of two nonoverlapping sets of 20 synonym pairs. These measures are: (*a*) the number of *popular responses* given by a speaker where a popular response is defined for each synonym pair as the most frequently occurring speaker response for that synonym pair, (*b*) the number of *unique responses* where a unique response is any speaker response to a synonym pair not given by any other speaker. The split-half reliability coefficients, based on the 272 speakers, were .041 and .042 ($p > .50$) for the popular and unique scores, respectively.

It is possible that the null hypothesis was not rejected because the individual-difference measures were based on too few word pairs. To check on this possibility, split-half reliability coefficients were also calculated from word-association data using the same number of items (20–20) and subjects (272). Word-association data from the same stimulus words as those in the speaker study were used. Split-half reliability coefficients based on the word-association data were .449 ($p < .001$) and .570 for popular and unique response scores, respectively. The sizable difference between these correlations and those based on the speakers' data suggest that a sufficient number of items was used to detect individual differences in communality among speakers if these differences existed.

A second study was designed to test the homogeneity assumption in a somewhat different way. In this study 50 subjects were each given 25 synonym pairs along with the standard speaker instructions. These 50 subjects were selected from the sample of subjects who had been in the word-association study described in a previous section. Their communality scores on the word-association test were used as a basis for selection; half of the selected subjects had high-communality scores (first quartile), and the other half had low-communality scores (fourth quartile). They were given the speaker task 6 months after their participation in the word-association study. The speaker-response frequencies from each of the two groups were compared. Response types given to each referent-nonreferent pair were grouped so as to yield an expected frequency of at least 5 in each cell of a groups (2) × response class matrix. A chi-square for each of the 25 referent-nonreferent pairs was then computed from this matrix of response frequencies. The sum of the 25 χ^2 values was 60.1 ($p < .30$, $df = 56$). Evidence from other work (Jenkins, 1960, p. 317) appears to rule out the possibility that statistical regression in word-association communality scores accounts for the absence of a difference between the two groups. The present result is in accord with the homogeneity assumption for the speaker task. It also raises the interesting question of why subjects who were consistently different in the word-association task were not also different in the speaker task. This question will be discussed in a later section on Intersubject Homogeneity in Word Association.

Word Associates and Speaker Responses

According to the theory—assumption *b* above—it would be expected that speaker responses to a referent are a subset of the word associates of that referent regardless of the nonreferent, of the details of the comparison stage, or of any changes in sampling probability that may occur when one or more sampled responses are rejected. However, a sample of word associates of a referent stimulus may not contain all of the obtained speaker responses to that referent even though the sampling assumptions are correct. This is evident from the fact that the proportion of speaker responses that also occur as word associates of *r* depends upon the sample sizes.

Therefore, it was necessary to determine the proportion of speaker responses that are likely to be present in a sample of word associates of a given size, assuming the theory is correct. This proportion was estimated empirically by using the proportion of *word associates* of *r* that are found at least once in an independent sample of word associates of *r*.

Data from 11 stimulus words were used in the analysis. These words were selected because they were on the Kent-Rosanoff list and, hence, were words with extensive normative data available from other investigators. Also they were referents in pairs of synonym words in the speaker study. For each stimulus word, we determined (*a*) the proportion of a sample of 272 word associates in our own word-association study which were found in samples of word associates collected by other investigators, (*b*) the proportion of a sample of 272 speaker responses present in the word-association norms of other investigators.[8]

Figure 4 summarizes the results. The curves are necessarily monotonic since they show the average proportion of samples of words (word associates or speaker responses) that were found in cumulative samples of word associates. That is, the leftmost point on any curve is the average proportion of responses found in a sample of 32 word associates given by Rutgers male undergraduates; the second point is the average proportion found in the Rutgers' sample of 32 *plus* a sample of 500 Minnesota male undergraduates, etc.

The bottom curve in Figure 4 was obtained by randomly pairing each of the stimulus words with another stimu-

---

[8] Word associates which would have violated the instructions to the speaker, for example, the nonreferent word, were deleted. For each stimulus word, the number of speaker responses used in the analysis was set equal to the number of word associates retained by randomly selecting this number of speaker responses from the total speaker sample of 272 responses.

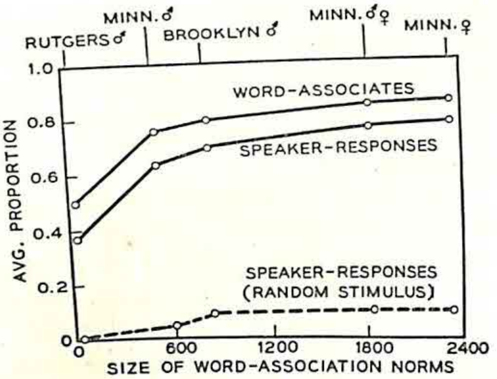

Fig. 4. The average proportion of word associates or speaker responses (ordinate) found in independent samples of word-association data (abscissa). (Based on 11 stimulus words from the Kent-Rosanoff list. Sources of word-association data are shown along the top of the graph—Rutgers males: Rothberg, 1965; Minnesota males: Palermo & Jenkins, 1964; Brooklyn males and females: Unpublished norms collected by S. J. Segal, 1960; Minnesota males and females: Russell & Jenkins, 1954; Minnesota females: Palermo & Jenkins, 1964.)

lus word from the set of 11, for example, ANGER, BEAUTIFUL, and counting the number of speaker responses to one (say ANGER) found in the word associates to the other (say BEAUTIFUL). This is essentially a check on the possibility that a substantial proportion of speaker responses can be found as word associates to any stimulus word if the sample of word associates is large enough. Apparently this is not true.

By contrast with the bottom curve, the two top curves are close together and may be converging. Furthermore, those speaker responses not found among the largest sample of word associates consisted almost exclusively of speaker responses that occurred once only. That is, the high-frequency speaker responses almost invariably showed up as word associates of the referent. It was concluded from these findings that, except for an inconsequential number of the I_r responses with the smallest $P_r(i)$ values, there are no responses in I_r which cannot be identified independently in a sizable sample of word associates.

Comparison Stage

It is difficult to verify, in detail, the assumptions concerning the comparison stage. This difficulty exists primarily because the sampling and comparison stages are so intertwined and because the sampling probability $P_r(i)$ "dominates" the behavior effects of the referential process.

To begin with we will demonstrate only the gross effects of the comparison stage. This will be done by first constructing a model consisting only of a sampling stage so that the nonreferent is assumed to have no effect on the emission of a speaker response. In such a one-stage model, the speaker response is essentially a random selection from the distribution of responses in I_r. It will then be shown that this one-stage model is inadequate and that these inadequacies suggest the existence of a comparison stage whose outcome depends on associative strengths $s_i(r)$ and $s_i(n)$.

Fitting a one-stage model to data is relatively straightforward. However, the unconditional probability, p_i, that the speaker emits i is not necessarily equal to $P_r(i)$ estimated from word-association norms since there are responses in the word-association norms that the speaker does not emit because he would violate the instructions not to give a response "that sounds like or is spelled like" either of the stimulus words. For these words it is assumed that $P_r(i) = 0$. For the remaining words, the relationship between p_i and $P_r(i)$ in the one-stage model is,

$$p_i = P_r(i)/\sum_{j \in I_r} P_r(j). \qquad [11]$$

Equation 11 is essentially a special case of Equation 2. That is, the assumption that a response that is sampled is invariably emitted is equivalent to setting $P_e(i|i)$ and $P_c(j|j)$ equal to 1 in Equation 2. In effect, the associative probability of a word that violates the instructions to the speaker is redistributed among the legitimate words in proportion to the $P_r(i)$ values of the latter. According to Equation 11 the relationship between p_i and $P_r(i)$ is linear.

The one-stage model summarized by Equation 11, in combination with word-association data, was used to predict the distribution of speaker responses from synonym pairs. Data from unrelated word pairs were treated separately for reasons which will be given later.

Figure 5 summarizes the relationship between the observed average proportion of speaker responses to synonym pairs and that predicted from word-association data using the one-stage model. Data from two speaker studies are plotted separately. One curve was obtained from an independent sample of 108 subjects used in a previous study (Rosenberg & Cohen, 1964); the second curve was obtained from the speaker study described in the present paper. Response types with similar predicted values were grouped to obtain the average values plotted in the figure.

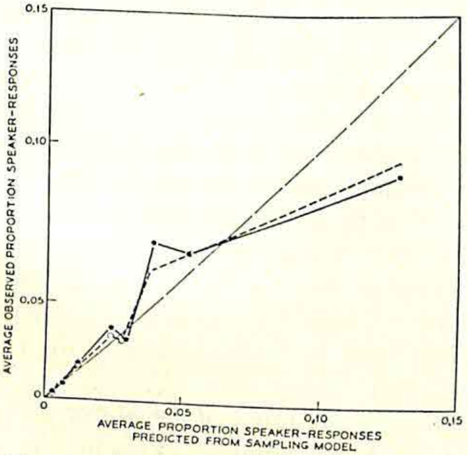

FIG. 5. Average observed frequency of speaker responses versus average frequency predicted from sampling (one-stage) model. (The diagonal is the set of points in which predicted frequency equals observed frequency. The solid curve is based on data from 40 synonym pairs—present speaker study. The dashed curve is based on data from 25 of these synonym pairs obtained in a previous study—Rosenberg & Cohen, 1964.)

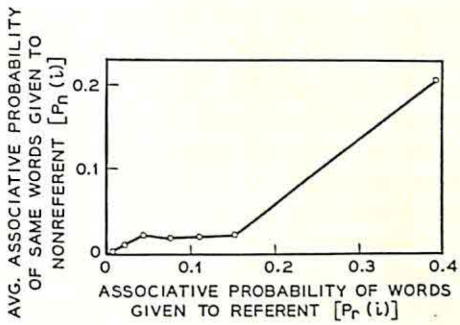

FIG. 6. The relation between the proportion of times that a response word occurred as a word associate of the referent $[P_r(i)]$ and the average proportion of times that the same response word occurred as a word associate of the nonreferent $[P_n(i)]$. (Based on 25 synonym pairs.)

If the sampling model were completely adequate, all data points would fall along the diagonal in the graph or, at least, would fall on a straight line. In both studies, a systematic deviation from this prediction occurs for the high-probability speaker responses where the observed frequency of speaker responses is considerably smaller than the predicted frequency. The question is how the comparison stage which was omitted for this analysis can account for this deviation from prediction.

A possible answer to this question was found by examining the associative probabilities to referent and nonreferent of these high-probability speaker responses. Figure 6 summarizes this relationship for 25 of the synonym pairs used in both speaker studies. The graph (which does not contain any speaker data) shows that the probability that a word associate of the referent will also be given as a word associate to the nonreferent is proportionately greater for high- than for low-probability associates of the referent. The effects of a comparison stage would thus be particularly noticeable for the high-probability word associates since their associative strengths to referent and nonreferent are both high. That is, the speaker will tend to reject the high-probability responses in I_r because of their similarity in associative strength to responses in the nonreferent set; low-probability responses, on the other hand, will be more likely to be emitted when they are sampled. The same effect was observed in a study by Krauss and Weinheimer[9] on the communication of colors. They demonstrated that when the hue of the nonreferent colors was similar to that of the referent, speakers used one of the less common (low-frequency) names to communicate the referent color. In dissimilar stimulus arrays the agreement among speakers in the names they used to refer to a color was extremely high. As we shall see, the latter finding corresponds to the behavior of our speakers when they are given unrelated word pairs.

For unrelated word pairs, the associative probability of any sampled response to the nonreferent is, for all practical purposes, zero. That is, a plot comparable to Figure 6 for an unrelated word pair would generally be a horizontal straight line with a close-to-zero intercept. Also, there are very few (if any) word associates of the referent that violate the instructions to the speaker. According to a sampling model, therefore, when the referent and nonreferent are unrelated, the distribution of speaker responses to a referent is expected to be essentially the same as that of the word associates to that referent.

However, according to the two-stage speaker model (sampling and comparison) the probability that the word i is a speaker response is greater than (less than) its probability, $P_r(i)$, as a word associate for high-probability (low-probability) word associates. An intuitive argument for this proposition is as follows. First, for unrelated word pairs the lower the associative strength of a sampled word to the referent, the closer this value will be to the associative strength of the sampled word to the nonreferent; the latter being the minimum possible value of associative strength. Hence, the speaker will probably judge remote associates as inadequate to communicate the referent to his listener. Second, in contrast to synonym nonreferents, unrelated nonreferents are not likely to have any high-probability associates in common with the referent. Hence, whenever a high-probability associate is sampled, the speaker will probably judge it as an adequate communication. The net result of these processes in the case of unrelated nonreferents, therefore, is to constrain the speaker towards clue words of higher communality than he might emit in standard free word association.

The formal proof of the above proposition is as follows. For unrelated word pairs we assume that the associative probability of a speaker response i to the nonreferent, $P_n(i)$, equals zero for all $i \epsilon I_r$. Thus, the associative strength $s_i(n) = A[P_n(i) + C]^B = AK$ where $A,B,C,$ and K are constants each > 0. From Equation 1 we obtain the conditional probability of emitting sampled response i:

---

[9] Unpublished manuscript, 1965.

$$P_e(i|i) = \frac{[P_r(i) + C]^B}{[P_r(i) + C]^B + K}. \qquad [12]$$

Equation 12 implies that $P_e(i|i)$ is a monotonically increasing function of $P_r(i)$. We now rewrite Equation 2 in the following form:

$$\frac{p_i}{P_r(i)} = \frac{P_e(i|i)}{\sum\limits_{j \epsilon I_r} P_r(j)P_e(j|j)}.$$

Since the sum in the denominator of the right-hand expression is a weighted average of all possible P_e $(i|\text{i})$ values, the equation is equal to a quantity less than 1 when $P_e(i/i)$ is at a minimum and to a quantity greater than 1 when P_e $(i|i)$ is at a maximum. In particular, $[p_i|P_r(i)] < 1$, that is, $p_i < P_r(i)$, when $P_e(i|i)$ is at a minimum. Furthermore, since $P_e(i|i)$ and $P_r(i)$ are monotonically related (Equation 12), $P_r(i)$ is also at a minimum when $P_e(i|i)$ is at a minimum so that $p_i < P_r(i)$ when $P_r(i)$ is at a minimum. By a similar argument it is found that $p_i > P_r(i)$ when $P_r(i)$ is at a maximum.

The empirical results are summarized in Figure 7. These results reveal that speaker-response types (dashed curve) occur with greater frequency than their frequency as word associates (solid curve) for the most common word associates of the referent stimulus; the reverse is true for the less common word associates. This observed relationship between the two curves is that predicted by the two-stage model.[10]

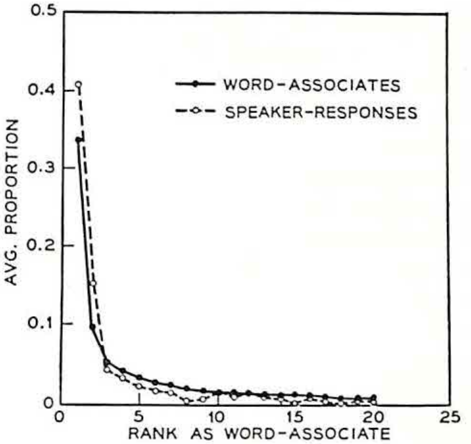

FIG. 7. Curve labeled "Word-Associates" shows the average rank-proportion relationship for responses given to 10 stimulus words. (Curve labeled "Speaker-Responses" shows the average rank-proportion relationship for the same responses given as speaker responses to the same stimulus words; the nonreferent in all stimulus word pairs was unrelated to the referent.)

By contrast, a one-stage model predicts that the two curves will coincide at all points. The χ^2 value comparing the frequencies of response types given as word associates and as speaker responses was 155.2 ($p < .00001$, $df = 20$). Thus, the speaker data from unrelated word pairs provide additional and somewhat different evidence for a comparison stage whose outcome depends on the associative strengths $s_i(r)$ and $s_i(n)$.

The data presented in Figure 7 also support the fundamental proposition that $P_r(i)$ can be estimated from the associative probability of i to r. From this proposition we would predict a strong monotonically increasing relationship between p_i and $P_r(i)$, providing $P_e(i|i)$ and $P_r(i)$ are either uncorrelated or are positively correlated. This monotonic relationship was also found in the data from synonym pairs although it is difficult to derive from the model for synonym pairs since the correlation between $P_e(i|i)$ and $P_r(i)$ is highly variable.

---

[10] A possible contaminating effect that was observed among the speaker responses was a heightened occurrence of synonyms as speaker responses when the two stimulus words were unrelated. This "synonym-response effect" may exaggerate the effect predicted from the two-stage model if synonyms are found more frequently among high- than among low-probability word associates. When the frequency of speaker responses (in a related type of communication task) was adjusted according to the correlation between frequency and synonymity, the effect predicted by the model was still highly significant (Rothberg, 1965). The synonymity between a stimulus word and its response was based on judges' ratings.

Goodness of Fit of the Two-Stage Model

The only serious obstacle to fitting the stationary Markov model (Equation 2) to the data is the fact that the parameters of the power function (relating associative strength to associative probability) are unknown. These unknown parameters are B and C in the function $s_i(x) = A[P_x(i) + C]^B$. (The parameter A cancels when the function is substituted into the Bradley-Terry model.)

One approach to the problem of estimating B and C is to assume that they are equal to the b and c, respectively, which enter into listener's comparison (Equation 6)—$\hat{b}$ and $\hat{c}$ would then be used as estimates of B and C. This is an attractive approach to the estimation problem since these parameter estimates are completely independent of the speakers' data.

Letting $\hat{B} = .700$ and $\hat{C} = .0018$ in the two-stage stationary model, we calculated a set of predicted p_i values for the speaker responses to each word pair. The chi-square values were then computed from the observed p_i values and the p_i values predicted from the model. The chi-square values were found to be of approximately the same size as those obtained from a comparison of obtained and predicted p_i values from the sampling (one-stage) model. In short, the parameter estimates obtained from the listeners' data are inadequate for use in the comparison stage of the speaker.

A second approach to the problem of estimating B and C is the use of an iterative least-squares procedure involving speakers' and word-association data. The sum:

$$\sum_i (p_i - \hat{p}_i)^2$$

was minimized with respect to B and C using an algorithm described by Marquardt (1963). This procedure did not yield estimates of B and C that resulted in any significant reduction of the chi-square values from those based on the parameters of listener's comparison. Since the properties of this iterative least-squares procedure (involving as it does a highly nonlinear model) are not well understood, we hesitate to conclude, without further research, that the comparison model used is inappropriate for the speaker. It may be the case, of course, that the Bradley-Terry model in combination with the power function is not appropriate for the speaker's comparison stage. For example, for sampled responses where $P_r(i) = P_n(i)$, it may be the case that $s_i(r) > s_i(n)$; that is, the conditional probability of emitting i may be greater than .5. That is, the speaker's decision to emit a response may be biased by the fact that he knows the referent. This type of bias cannot be accounted for by any combination of values of B and C. A possible answer to this question may be furnished by a study in which the sampling of a response is done for the speaker by the experimenter, and the speaker has the choice of using the response as a speaker response or rejecting it as inadequate.

INTERSUBJECT HOMOGENEITY IN WORD ASSOCIATION

Because of the intimate relationship between I_r and word-association norms, it is natural to extend, if possible, the assumption of intersubject homogeneity to the word-association situation. One of the immediate consequences of the intersubject homogeneity assumption in the word-association situation is that the response protocol (one response to each of a set of stimulus words) for any individual subject will contain as many popular and unique responses as that of any other randomly selected subject. The facts, however, clearly contradict this proposition. Highly reliable differences among subjects in their communality scores have been demonstrated in the results of other investigators (Fosmire & Tryk, 1963; Jenkins, 1960) as well as in our own word-association data. (In our data the split-half reliability coefficient based on responses to 152 stimulus words was .82.) In other words, different subjects contribute disproportionately to different parts of an average rank-frequency distribution of word associates.

How can these facts be reconciled with the findings that group word-association norms do provide satisfactory estimates of I_r and that intersubject homogeneity in the speaker task is a tenable asssumption? It may be the case that underlying response sets in the word-association situation vary from subject to subject, but group norms somehow estimate I_r accurately, and these I_r sets do not differ from speaker to speaker. This ad-hoc assumption is clearly not as attractive as the assumption of intersubject homogeneity in both word-association and speaker tasks. In addition, the question of intersubject homogeneity in word association also arises in connection with the use of associative strength and associative overlap based on word-association norms in a large variety of studies of verbal learning and mediation (Cofer, 1961; Deese, 1962).

The highly reliable individual differences in communality scores obtained from word-association responses can, however, be reconciled with a homogeneity assumption by postulating a two-stage process in the word-association task. According to this conception, the subject in the word-association task first samples a response from a hypothetical set of responses and then compares certain properties of this sampled response with criteria supplied by the experimental instructions (as in controlled word association) or supplied by the subject himself (as in free word association). The subject either emits the sampled response or rejects it depending on how closely the sampled response approximates the assigned or self-instructed criteria. With such a two-stage model the assumption can be retained that all subjects from a judiciously chosen population sample from substantially the same response set, but that the subjects may vary in their self-instructed criteria.

A number of investigators have noted the existence of several self-instructed criteria (of which the subject may not be aware). In an extensive study of this question, Moran, Mefferd, and Kimble (1964) used a factor-analytic technique and identified three distinct self-instructed criteria which they term "idiodynamic sets." Moran et al. also present evidence to the effect that when the popular responses to a stimulus word are congruent with the subject's idiodynamic set, the subject contributes disproportionately to the frequency of these responses and makes fewer "mediation faults," for example, giving multiword responses, failing to respond within the time limit, giving remote associates such as *box* to FAITH, etc. These findings are consistent with a two-stage model for word-association behavior in which the idiodynamic set is operative during the comparison stage only. From the results of the speaker studies reported in the previous section, it is concluded that these idiodynamic sets are suppressed in the communication task in favor of a uniform communication set.

If the two-stage model does hold for free word association, however, there is another dilemma. How can free-word-association norms be used to estimate sampling probabilities from I_r if these norms have been biased by a variety of subject-controlled criteria during comparison in free word association? One possibility is that the effect of the sampling stage in word association is strong enough to prevent these biases (in the comparison stage) from seriously distorting the sampling probability of most responses in the underlying set of associates. The results of our experimental analyses concerning the sampling stage in the speaker's referential process suggested that this stage is the dominant one. An alternative hypothesis is that these idiodynamic sets "cancel" each other so that group norms yield adequate estimates of the sampling probabilities. Of course, these hypotheses require further empirical study. For the moment, the justification for using free-word-association norms in tests of the speaker model is essentially the same as it is in studies of verbal learning and mediation—an heuristic one.

RECOGNITION AND RECALL PROCESSES

The essential operational difference between recall and recognition is that in recall the subject is required to reproduce the items on a previously presented training list, while in recognition the subject is required only to judge whether items, produced for him by the experimenter, were on the training list. It is suggested that performance on recall tasks is determined by a two-stage process formally similar to the present speaker model, and recognition by a one-stage process similar to the present listener model.

According to this conception, in recalling an item, a subject first samples an item from an underlying set that consists of a list and an extralist subset. He then goes through a comparison stage that determines the probability that the sampled item will be emitted. In the comparison stage of the recall process, the subject, who must decide whether or not the sampled item is in the list subset, may do this by comparing the associative strength of his sampled response with a constant representing a threshold value of associative strength. That is, the probability that an item will be emitted, given that it has been sampled, equals $t_i/(t_i + k)$, where t_i is the subjective probability (associative strength) that the sampled item is a list member, and k is some minimal value of associative strength below which the subject will not accept the item as a list member. In a recognition test, only the comparison stage is entailed.

The model of recall *learning* recently proposed by Kintsch and Morris (1965) is related to the two-stage model described in the present paper. They hypothesize that recall learning is composed of a "recognition stage and a 2nd stage where the response becomes available in the absence of the stimulus [p. 200]." In effect, in learning there is a change in sampling probability of an item and a change in the subjective associative strength of the item above the threshold value k. Thus recognition may be superior to recall for at least two reasons: (*a*) In recognition the sampling stage is eliminated as a variable affecting performance; (*b*) the probability of correctly recognizing an item as a list member increases faster than its sampling probability. The latter effect corresponds to a finding observed in our data: A speaker response often exists which would result in a low proportion of listener errors, but is rarely given by speakers because it is a remote associate to the referent; that is, it has a low sampling probability.

The advantages of recognition over recall can be considerably reduced by manipulations that raise the sampling probabilities of list members and proportionately lower those of the extralist members of the underlying set, for example, overlearning or the use of contextual cues. This could make the recall task functionally equivalent to a recognition task with respect to item availability. A study by Lachman and Field (1965), for example, showed that recall scores did reach, and even surpass, recognition scores when the list of items was presented a sufficient number of times prior to the retention tests. The actual superiority attained by the recall scores in Lachman and Field's (1965, p. 25) study, after 16 or more training trials, can be attributed to the fact that the training list was presented in a serial order that had a "high approximation to English (i.e., prose)." On the recall tests, this had the effect of further raising not only the sampling probabilities but also the associative strengths of the items, since the subjects were free to use the sequential dependences among the items as contextual cues during recall. On the recognition tests, however, subjects were given list and nonlist items in random order. Hence, the sequential dependences among the items could not provide the same increments in associative strength on the recognition as on the recall tests, since the items were always sampled for the subject out of context.

In multiple-choice recognition tests of the type analyzed by Murdock (1963), the present listener model appears to be directly applicable. The subject is in essentially the same situation as the listener in the word-communication task; he must choose a correct item from two or more alternatives. The listener-comparison model is appropriate to the job of predicting the proportion of correct choices in such situations provided that information is available concerning the associative strengths of the individual alternatives.

Data relevant to this proposition were obtained in a probabilistic paired-associate study by Donner (1965) in which retention was measured with a two-alternative recognition test. In the training phase, Donner showed subjects ($N = 100$) a series of 160 letter-number pairings flashed on a projection screen at a .7 second/.7 second on-off rate. Each pairing consisted of one of four letters (F, M, P, W) and one of four digits (3, 6, 8, 9). The subjects were told to "learn the various combinations that appear. The same letter may appear with a different number [p. 3]." The purpose of this procedure was to generate a set of associative probabilities of each of the response terms (the digits) to each of the stimulus terms (the letters), and then to use these probabilities as the word-association normative probabilities were used in the earlier listener study.

In the recognition test, the subjects were given a series of triples. Each triple consisted of two of the four stimulus letters and one of the four response digits. Subjects were required to choose the letter that "best goes with the number on the basis of what you have learned [p. 9]." The triples used in these tests were selected to permit assessment of subjects' choices over a wide range of differences in the associative probabilities of the response term to each of the stimulus terms.

The results were fitted to the listener model (Figure 8) using as estimates of $P_x(i)$ and $P_y(i)$ the proportion of times a response i was paired with the stimuli x and y, respectively, during the

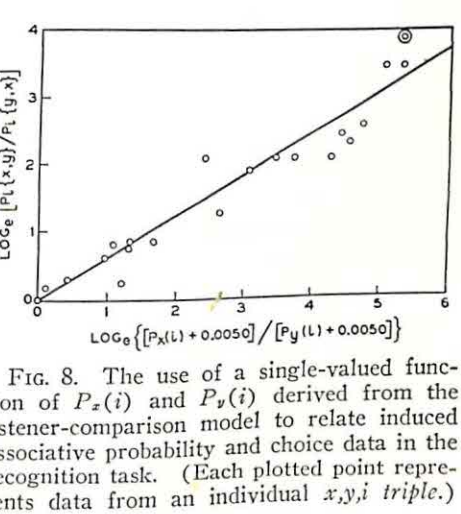

FIG. 8. The use of a single-valued function of $P_x(i)$ and $P_y(i)$ derived from the listener-comparison model to relate induced associative probability and choice data in the recognition task. (Each plotted point represents data from an individual x,y,i *triple*.)

training phase. The proportion of subjects choosing x in the recognition test was used as an estimate of $P_i\{x,y\}$. The parameters b and c in the model were estimated by the iterative least-squares procedure used in the listener study. The slope $\hat{b}$ of the fitted line in Figure 8 is .623. These parameter estimates probably depend, in part, on the number of training trials, but are similar in order of magnitude to those found in the listener study with word-association data. As in the listener study, special cases of the power function, that is, $b = 1$ or $c = 0$, resulted in considerably poorer fits. Donner (1965) also showed that the simple difference, $P_x(i) - P_y(i)$, was not adequate to predict recognition data.

The models outlined in this paper were constructed to account for speakers' and listeners' responses in referent communication; the Donner study and the test of the choice axiom with listener data suggest that these models may also have promise for the development of new approaches to some of the more conventional problems in verbal learning and performance. Rote-learning techniques might, in turn, permit more definitive tests of the speaker model than have so far proven possible. That is, such techniques might permit more precise specification and control of the sets from which speakers' responses are sampled, and the sampling probabilities and associative strengths of the responses in these sets.

REFERENCES

ADAMS, E., & MESSICK, S. An axiomatization of Thurstone's successive intervals and paired comparisons scaling models. Technical Report No. 12, Stanford University, Applied Mathematics and Statistics Laboratory, 1957.

BRADLEY, R. A., & TERRY, M. E. The rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika,* 1952, **39**, 324–345.

BROWN, R. W. *Words and things.* Glencoe, Ill.: Free Press, 1958.

COFER, C. N. (Ed.) *Verbal learning and verbal behavior.* New York: McGraw-Hill, 1961.

DEESE, J. On the structure of associative meaning. *Psychological Review,* 1962, **69**, 161–175.

DONNER, L. Choice behavior in a verbal recognition task as a function of induced associative strength. Unpublished master's thesis, Rutgers University, 1965.

FOSMIRE, F. R., & TRYK, H. E. Word association: Common and original response. *Science,* 1963, **139**, 415–416.

JAKOBSON, R. Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language.* New York: Wiley, 1960. Pp. 350–377.

JENKINS, J. J. Communality of association as an indicator of more general patterns of verbal behavior. In T. A. Sebeok (Ed.), *Style in language.* New York: Wiley, 1960. Pp. 307–329.

KANTOR, J. R. *An objective psychology of grammar.* Bloomington, Ind.: Principia Press, 1936.

KEMENY, J. G., & SNELL, J. L. *Finite Markov chains.* New York: Van Nostrand, 1960.

KINTSCH, W., & MORRIS, C. J. Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology,* 1965, **69**, 200–206.

KRAUSS, R. M., & WEINHEIMER, S. Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science,* 1964, **1**, 113–114.

LACHMAN, R., & FIELD, W. H. Recognition and recall of verbal material as a function of degree of training. *Psychonomic Science,* 1965, **2**, 225–226.

LANTZ, D., & STEFFLRE, V. Language and cognition revisited. *Journal of Abnormal and Social Psychology,* 1964, **69**, 472–481.

LUCE, R. D. *Individual choice behavior.* New York: Wiley, 1959.

MACLAY, H. A descriptive approach to communication. In N. F. Washburne (Ed.), *Decisions, values and groups.* New York: Pergamon Press, 1962. Pp. 201–226.

MACLAY, H., & NEWMAN, S. Two variables affecting the message in communication. In D. Willner (Ed.), *Decisions, values and groups.* New York: Pergamon Press, 1960. Pp. 218–228.

MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters.

Journal of the Society for Industrial and Applied Mathematics, 1963, **11**, 431–441.

MARSHALL, G. R., & COFER, C. N. Associative indices as measures of word relatedness: A summary and comparison of ten methods. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 408–421.

MEAD, G. H. *Mind, self and society*. Chicago: University of Chicago Press, 1934.

MORAN, L. J., MEFFERD, R. B., JR., & KIMBLE, J. P., JR. Idiodynamic sets in word association. *Psychological Monographs*, 1964, **78**(2, Whole No. 579).

MORRIS, C. *Signs, language and behavior*. New York: Prentice-Hall, 1946.

MURDOCK, B. B., JR. An analysis of the recognition process. In C. N. Cofer & B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes*. New York: McGraw-Hill, 1963. Pp. 10–22.

OGDEN, C. K., & RICHARDS, I. A. *The meaning of meaning*. New York: Harcourt, Brace, 1923.

OSGOOD, C. E., & SEBEOK, T. A. (Eds.) *Psycholinguistics: A survey of theory and research problems*. Bloomington: Indiana University Press, 1954. (Publ. in Anthropology and Linguistics, Mem. 10.)

PALERMO, D., & JENKINS, J. J. *Word association norms*. Minneapolis: University of Minnesota Press, 1964.

QUINE, W. V. O. *Word and object*. Cambridge: Massachusetts Institute of Technology Press, 1960.

ROSENBERG, S., & COHEN, B. D. Speakers' and listeners' processes in a word-communication task. *Science*, 1964, **145**, 1201–1203.

ROTHBERG, M. The effect of "social" instructions on word-association behavior. Unpublished master's thesis, Rutgers University, 1965.

RUSSELL, B. *The analysis of mind*. New York: Macmillan, 1921.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Technical Report No. 11, 1954, University of Minnesota, Contract No. N8-onr-66216, Office of Naval Research.

SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

SNEDECOR, G. W. *Statistical methods*. Ames: Iowa State College Press, 1946.

STEVENS, S. S. On the psychophysical law. *Psychological Review*, 1957, **64**, 153–181.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

WERNER, H., & KAPLAN, B. *Symbol formation*. New York: Wiley, 1963.

(Received January 18, 1965)

Psychological Review
1966, Vol. 73, No. 3, 232–240

INHIBITION OF THE UNCONDITIONED RESPONSE IN CLASSICAL CONDITIONING[1]

H. D. KIMMEL[2]

University of Florida

The classical conditioning process is characterized as a "dynamically" gradual transition from an initial excitatory state, in which the CS produces a short-latency CR, to an increasingly inhibitory condition, in which both the CR and UCR are attenuated and the former is greatly delayed, depending upon the CS–UCS interval used during acquisition. The highest point in the CR magnitude function is assumed to occur in fewer trials when long interstimulus intervals are used and in more trials when shorter intervals are used. UCR diminution is shown to depend upon an associative process by the device of omitting the CS and noting the increase in UCR strength thus produced. This increment has been found to be a function of the number of prior paired conditioning trials, the CS–UCS interval (in conventional fashion, with a peak at about .5 sec.), and whether a delayed or trace conditioning paradigm is used. It is suggested that the development of inhibition of delay, upon the basis of which these effects are assumed to occur, may be similar to differential or selective reinforcement (of an instrumental type) of late as compared with early responses. When the CR is made instrumental in shock avoidance, CR attenuation and UCR recovery are not observed. A study of the relationship between GSR amplitude and judged intensity of electric shock failed to provide support for the notion that this reinforcement may be due to reduced noxiousness of the UCS when late responses occur, but the issue is not closed by these results.

The purpose of this paper is to bring together under a relatively parsimonious theoretical mantle a variety of empirical observations, most of which have only recently become part of our classical conditioning literature. The basic proposition elaborated below is that classical conditioning, rather than a "dumb" or senseless kind of learning (Tolman, 1938) or a mere artifact of instrumental learning (Smith, 1954), is, in fact, a quite flexible and eminently adaptive process. Since the basic argument involves, in some ways, a "marriage" of classical and instrumental conditioning concepts, it should come as no surprise that "something old and something new" are present at the ceremony. The "something old" is Pavlov's (1927) concept of *inhibition of delay*. The "something new" concerns the effect of such inhibition on the magnitude of the unconditioned response (UCR).

In classical salivary conditioning the adaptive function of inhibition of delay seems quite apparent. It is surely more economical to the dog to salivate late in the interstimulus interval, just before the food or acid unconditioned stimulus (UCS) is to be delivered, than to do so at the very onset of the conditioned stimulus (CS), long before the secretion can be of use in the mastication and digestion of the food or in the dilution of the acid. It is less apparent that delay of, say, a conditioned galvanic skin response (GSR) or eye blink could be of any adaptive value, yet this is what we will suggest. For expository purposes we will deal first with studies in which the CS has been shown to have an inhibitory influence upon the conditioned response (CR) and then with studies in which the UCR is similarly inhibited by the CS. As will be brought out below, these effects are better thought of as manifestations of a unitary process influencing both responses, the present distinction being one of convenience only.

---

[1] Done under Grant MH-06060-02 from the National Institute of Mental Health.

[2] Now at Ohio University.

Inhibition of the CR by the CS

The phenomenon of inhibition of delay, itself, exemplifies the manner in which the CS gradually becomes capable of attenuating and delaying the CR it had previously been conditioned to elicit. In salivary conditioning, Pavlov (1927), of course, has amply demonstrated the degree to which such inhibition accrues under conditions of increasingly long delays of the UCS, and, in addition, he has shown how an exteroceptive stimulus presented during the latent interval acts to disinhibit this inhibition and bring forth an immediate and sometimes increased salivary CR. Wenger (1936) and Rodnick (1937a) have reported more or less similar results with the human GSR and eyelid reflex. Rodnick's (1937a) study showed, further, that greater inhibition of delay was present using a "delayed" paradigm than with a "trace" paradigm, tying the process even more closely to the *presence* of the CS at the moment the UCS occurs rather than to its mere onset.

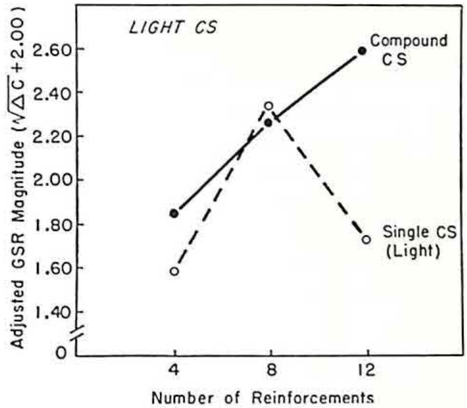

FIG. 1. GSR response to trained CS and to CS plus extra stimulus (compound) following 4, 8, and 12 reinforcements of a single stimulus.

The degree to which an added stimulus produces an increment in GSR amplitude has been shown to depend upon the number of CS–UCS pairings prior to the introduction of the extra stimulus (Kimmel & Fowler, 1961; Kimmel & Greene, 1964). Figure 1 summarizes a portion of the results of the former study. In Figure 1 it can be seen that the average GSR to the combination of the visual CS and an added auditory stimulus significantly exceeded the average GSR to the visual CS alone only following 12 CS–UCS pairings, but not following either 4 or 8 pairings. The response to the trained CS alone rose for eight trials and then fell. This effect is even more striking when a greater number of paired trials are used before introducing the novel stimulus (Kimmel & Greene, 1964). In that study it was shown that the response to the novel stimulus continued to rise in negatively accelerated fashion throughout 50 paired trials, while the response to the CS alone first rose and then dropped.

The Kimmel and Greene study also showed that the time during the CS–UCS interval at which the extra stimulus is delivered is a variable which greatly influences the ultimate response pattern observed at different stages of conditioning. In that study a 7.5-second interstimulus interval was used, and the extra stimulus was delivered at five different points in the interval—at its onset, after one-fourth, one-half, or three-fourths of its duration, or at its termination. Kimmel and Greene found that the time of onset of the extra stimulus during the CS–UCS interval mattered little prior to conditioning, that after a couple of conditioning trials the response to the novel stimulus was greatest at the beginning of the CS–UCS interval, and that late in conditioning the response to the novel stimulus was greatest just before the delivery of the UCS. They interpreted their results to mean that, during conditioning, inhibition of delay produces a gradual shift in the focus of excitation within the CS–UCS interval, from its beginning to its end. The ultimate consequence of inhibition of delay, as Pavlov himself noted, may be the total inhibition of the CR during the interstimulus interval.

It is of some interest to note that this form of inhibition, that is, initiated and controlled by the CS during the interstimulus interval, may result in organism-wide tranquility. Rodnick (1937b), for example, found that the eyelid reflex, conditioned to a different CS previously, was depressed when measured during the long delay interval of the CR currently being conditioned to a new CS. Spence and Runquist (1958) showed that the unconditioned eye blink to a weak air puff was depressed when elicited after a CS previously paired with electric shock had been on for $\frac{1}{2}$ second. Tranquility during conditioning has also been found to lead to drowsiness and even sleep by Pavlov (1927).

Pavlov recognized that every CS, even in the presence of reinforcement, regularly acquires inhibitory as well as excitatory properties. Indeed, he wrote: "The function performed by the unconditioned reflex after the conditioned reflex has become established is merely to retard the development of inhibition [1927, p. 234]." That this is true in GSR conditioning is relatively well known (Kimmel, 1964). Several studies have shown that the course of the curve of conditioned GSR amplitude during reinforcement is first upward, for varying numbers of trials, and then downward, until eventually it may fall below even the level observed before conditioning (Grings, Lockhart, & Dameron, 1963; Kimmel, 1959; Meryman, 1953). The number of trials required for the highest point in the curve to be reached appears to be inversely related to the CS–UCS interval employed. Meryman (1953) used an interval of 19 seconds and found a peak in the CR amplitude curve at about two or three trials. Kimmel (1959) used a 4-second interval and observed a peak at six to eight trials. Grings et al. (1963) ran two groups with intervals of .5 second and 5.0 seconds and found a peak in the 5.0-seconds condition at about five trials, while the CR amplitude curve was still rising in the .5-second condition after eight trials. Support for this generalization may be found even in eyelid-conditioning data. Ost (1960) found that the highest point in the percentage of CR curve was reached after only 30 conditioning trials with a 1.0-second interval, but that the curve was still rising after 50 reinforcements with a .5-second interval.

These studies support the notion that inhibition of the CR associated with the CS develops more readily and in fewer trials when long rather than short interstimulus intervals are used. Logical extension of this relationship leads to the rather obvious prediction than an infinite interstimulus interval would result in a peak in the CR amplitude curve at zero trials (conditioning would not occur at all). At the other extreme, a zero interval would require an infinite number of trials for the peak of the curve to be reached, which is in accord with Pavlov's assertion that the UCS (here delivered simultaneously with the CS) acts to overcome the inhibition normally being attached to the CS. Thus it is not surprising to note that simultaneous conditioning fares poorly (McAllister, 1953; White & Schlosberg, 1952).

Inhibition of the UCR by the CS

The extent to which the amplitude of the UCR may be attenuated by the CS during classical conditioning is best demonstrated by the technique of omitting the CS after several pairings of it with the UCS and presenting the UCS alone. Kimble (see Hilgard & Marquis, 1961) has described an eyelid-conditioning experiment that represents the first published report of this procedure. In a study by Kimble and Dufort (1956) on the effect of interpolated UCS presentations on eyelid conditioning, a series of 20 CS–UCS trials was followed by 20 presentations of the UCS alone and, then, 20 additional CS–UCS trials. An incidental observation in this study was that the size of the UCR, which had diminished gradually during the first 20 CS–UCS trials (an effect previously noted by Hilgard & Marquis, 1961; attributed by them to reflex habituation), increased abruptly on Trial 21 when the UCS was not preceeded by the CS. The UCR remained at a high level until Trial 41 when the CS was again presented just before the UCS, and the UCR was again attenuated. This reduction in size of the UCR in the presence of the CS was not attributable to its having been elicited during a refractory period following the CR, since the reduction did not appear to be correlated with the presence of the CRs. Indeed, the reverse was more likely, since individuals showing the fewest CRs were the ones who evidenced the most reduction in the UCR.

Kimble and Ost (1961) reasoned that the reduction in UCR amplitude was due to some form of learned inhibition. To check on this, they studied the relationship between UCR diminution and recovery and the CS–UCS interval that had been used prior to omitting the CS. They found that the greatest degree of recovery of the UCR produced by omitting the CS occurred in a group of subjects conditioned with an interstimulus interval of .5 second. As a matter of fact, their curve of the relationship between the CS–UCS interval and degree of UCR recovery was almost indistinguishable from typical curves of the so-called interstimulus-interval function in conventional conditioning studies (McAllister, 1953; Moeller, 1954; Prokasy, Fawcett, & Hall, 1962). Kimble and Ost concluded that a conditioned inhibitory process resembling Pavlov's inhibition of delay must have been responsible for the reduction in UCR amplitude in the presence of the CS.

UCR reduction in the presence of the CS has also been found in salivary conditioning (Kupalov, 1961). In his description of "pathological" irradiation of inhibition in one of his dogs, Kupalov noted that the UCR was completely eliminated during an extended series of reinforced delayed conditioning trials. He used the term pathological because the dog eventually even failed to secrete saliva during normal feeding without any CS at all.

Diminution of the UCR in GSR conditioning appears to be equally demonstrable. Lykken (1959) has reported that the amplitude of the GSR to an electric-shock UCS which was repeated every 30 seconds was sharply reduced when a tone was presented 5 seconds before the shock. His interpretation of this observation involved the cognitive-perceptual concept of *preception,* the subjects presumably being better able to estimate the time of delivery of the shock following a 5-second warning than following a 30-second warning. Viewed in this way the UCR diminution and recovery phenomenon in GSR conditioning would seem to be a close relative of what Grings (1960) and Kimmel (1960) have referred to as a *perceptual disparity response,* except that the latter phenomenon was observed in situations in which the CS was changed rather than omitted or the UCS was changed and the CS kept intact. Evidence presented below will be used to clarify the cognitive-associative question and provide support for one position over the other.

Pursuing the lead suggested by the Kimble and Ost (1961) study, Kimmel and Pennypacker (1962) gave 4, 8, or 16 CS–UCS presentations to different groups of subjects prior to presenting the electric-shock UCS alone. They found a significant decrement in UCR amplitude (GSR) during reinforcement followed by a large increment when the CS was omitted. The amount of increase in UCR amplitude produced by omission of the CS was linearly related to the logarithm of the number of prior reinforcements; that is, a negatively accelerated growth function was obtained. Following this, Kimmel and Pennypacker,[3] in an unpublished experiment, found evidence that the degree of recovery of the UCR due to omission of the CS followed the conventional interstimulus-interval function, as had previously been found by Kimble and Ost (1961), although the statistical significance of this result was only marginal (p <.10). The overall increase in UCR amplitude following omission of the CS was, however, highly significant.

Since the results of these studies provided further support for the proposition that diminution of the UCR during classical conditioning is, itself, conditioned, Morrow (in press) studied the possibility that it would be extinguishable in the conventional sense. He gave three groups of subjects 3, 6, or 12 CS-only extinction trials following paired CS–UCS trials and, after this extinction, presented a paired CS–UCS trial. All three groups made larger UCRs following extinction than they had before extinction (i.e., the conditioned UCR diminution had been extinguished as hypothesized), a result not found in three control groups that had received the CS and UCS in an unpaired series of presentations prior to extinction and the CS–UCS test trial. No significant differences in amount of extinction of UCR diminution were found to be associated with the different numbers of extinction trials, a finding that led Morrow (in press) to speculate that extinction of this type of inhibition may be more rapid than is typical.

The extent to which the CS controls diminution of the UCR in GSR conditioning was studied by Baxter (1966). Half of his conditioning subjects received a 5-second tone paired with electric shock (delayed paradigm), while the other half received a 1-second tone whose onset preceeded the shock by 5 seconds (trace paradigm). Thus the CS–UCS interval was 5 seconds in both groups. Two control groups were run in which the 1- or 5-second tone and shock were delivered in an unpaired sequence. Test trials on the UCS alone were presented following either conditioning or the control treatment. While diminution of the UCR was observed in both of the paired groups, it occurred to a significantly greater extent in the delayed (5-second CS) conditioning group than in the trace (1-second CS) group. Similarly, omission of the CS led to greater recovery of the UCR in the delayed group than in the trace group. The control groups showed only slight UCR reduction and no recovery on test trials. These results support the conclusion that the actual physical presence of the CS leads to greater inhibition of the UCR than occurs in the presence of the "trace" of a CS. The slight reduction in UCR size that was observed in the control groups was attributed to habituation rather than an associative mechanism.

---

[3] Done while the author was a visiting assistant professor at Duke University, 1961.

Baxter's (1966) results are relevant also to the question of whether UCR diminution is a simple conditioning phenomenon or would be better understood in perceptual-cognitive terms. If the reduction in the size of the UCR results from the fact that the subject comes to "know" that the UCS follows the CS and, consequently, "prepares" to receive it when the CS is presented, it would be reasonable to assume that this preparation involves a temporal parameter, namely, the time when the shock will be presented following the CS. But in Baxter's experiment both conditioning groups had a 5-second warning period from the onset of the CS to the onset of the UCS and both, under this reasoning, should have developed similar preparatory sets. If any perceptual advantage existed in the situation, it should have favored the trace conditioning group, since it was probably easier to estimate the 4 seconds from CS offset to UCS onset in the trace condition than it was to estimate the 5 seconds from CS onset to UCS onset in the delayed condition. It seems more likely that the diminution phenomenon depends more upon the presence of CS initiated and controlled inhibition than upon the subject's preparation to receive the shock.

A perceptual approach would "explain" the increment in the size of the UCR produced by omission of the CS as a manifestation of the subject's "surprise" at receiving the UCS alone. Again, however, Baxter's (1966) results do not support this position. On what basis would one expect the delayed conditioning procedure to result in greater surprise at receiving an unpaired shock than occurred in the trace conditioning group? Presumably, the argument would be made that the trace paradigm is somewhat more similar to no pairing at all than is the delayed paradigm. This argument, however, since it places its emphasis upon the contiguity of CS offset and UCS onset, is a rather different perceptual theory than that usually proposed to handle the acquisition of conditioning, as exemplified in the preceeding paragraph. In addition to the groups already described above, Morrow (in press) ran an additional control group that received 12 unpaired shocks following conditioning, instead of the extinction trials of CS alone. As had previously been reported for the eyelid reflex by Kimble (see Hilgard & Marquis, 1961), reintroduction of CS–UCS pairing following UCS-only trials, a change in procedure no less surprising than omission of the CS, resulted in a *drop* in the size of the UCR. On the basis of the results obtained in these studies, it would appear that a simple conditioning model provides a more adequate theoretical format than does the less parsimonious, albeit more "human," cognitive-perceptual interpretation.

Instrumental Features of UCR Diminution

We have assumed that reduction in the size of the UCR in classical aversive conditioning is a manifestation of inhibition of delay and is thus under the control of the CS. When the CS is omitted, the inhibition is absent, and the UCR is larger. Two questions are suggested by this argument. First, is it possible that inhibition of delay develops as a result of differential reinforcement of late as compared with early CRs? Second, does the assumed differential reinforcement in question result from the fact that smaller UCRs reflect reduced noxiousness in the subject's experience? Empirical answers to both of these questions have recently been sought.

What would we expect to happen to the UCR in a GSR conditioning experiment in which the CR is made instrumental (i.e., leads to avoidance of the shock)? If inhibition of delay is the result of differential reinforcement of late as compared with early responding in the CS–UCS interval, we would have to predict that UCR recovery when the CS is omitted would not occur appreciably in an avoidance situation, since it would be more adaptive to make an early CR and avoid the shock than it would to develop inhibition of delay and receive a possibly less noxious shock.

Kimmel (1965) has recently described a GSR conditioning experiment in which the subjects were run in yoked pairs such that whenever the avoidance subject of a pair made a CR of criterion magnitude both subjects received no shock on that trial. Half of the subjects were run with a 5-second CS and half with a 1-second CS, but all subjects had a 5-second CS–UCS interval. Following the acquisition trials, a test trial on the UCS alone was given to all subjects. The avoidance groups showed a slight decrement in the size of the UCR when the UCS was delivered alone following conditioning, while both the delayed and the trace yoked classical groups showed the expected increment, although the difference between avoidance and yoked control groups was statistically significant only in the delayed paradigm. These results are in accord with the assumption that the CS acquires its inhibitory properties only in a classical conditioning situation, but not when the CR is made instrumental. When an early CR results in shock avoidance, omission of the CS at the end of the training does not result in an increment in the size of the UCR. Also pertinent to this argument is the recent observation (Kimmel & Baxter, 1964) that inhibition of the CR also fails to occur when an avoidance paradigm is used in GSR conditioning.

Does the fact that a smaller GSR is made to the shock when it is paired with the CS than when it is presented alone mean that the shock is perceived as less intense? Kimmel and Schultz (1964) had their subjects judge the intensity of electric shocks ranging from .5–3.5 milliamperes at the same time that GSRs were measured. Both judged intensity and GSR magnitude were significantly positively related to shock intensity, but the correlation between GSR magnitude and judgments was found to be significantly positive in only one of the nine subjects that were observed, when the effect of shock intensity was partialed out statistically. Since a significant negative correlation was found in the data of one other subject, the results of this study surely do not provide support for the assumption that the perceived intensity of the shock holds the explanatory key to a differential reinforcement interpretation of UCR diminution and recovery.

Summary and Conclusions

The studies reviewed in this paper support the conception that the acquisition of classical conditioning involves a distinctly negative component side by side with its more well-known positive characteristics. After the formation of an initial association of an excitatory nature, the CS gradually begins to develop inhibitory properties that result in attenuation and delay of the CR and even its ultimate total loss under continued reinforcement, as well as diminution of the UCR in the presence of the CS. The latter is revealed particularly by the technique of omitting the CS and comparing the resulting UCR with UCRs observed (both before and after) when the CS is present. UCR recovery has been shown to vary as a function of the CS–UCS interval used during training (Kimble & Ost, 1961), the number of reinforced trials prior to omitting the CS (Kimmel & Pennypacker, 1962), and whether a delayed or trace conditioning paradigm was used prior to CS omission (Baxter, 1966). The proposition that UCR diminution and recovery depend upon a conditioning or associative process is further supported by the observation that it does not occur appreciably in an unpaired sequence of CSs and UCSs (Baxter, 1966; Morrow, in press), and by the finding that standard extinction procedures result in its loss (Morrow, in press). Several features of these results do not accord with the notion that they are due to a cognitive-perceptual type of process, although further research on this issue is surely not precluded.

The possibility that CS-controlled inhibition is related to an instrumental differential-reinforcement process receives some support from the fact that CR inhibition does not occur in an avoidance conditioning situation and that UCR recovery when the CS is omitted is similarly absent in avoidance GSR conditioning while present in a yoked classical control condition. That the mechanism of such reinforcement is to be found in the perceived noxiousness of the UCS, while not supported in one study of the relationship between GSR magnitude and judged intensity of electric shock, must still be considered a possibility and should be the object of future research in this area.

If they do nothing else, the results of this line of investigation can be taken to mean that a new view of classical conditioning deserves serious consideration. Rather than a purposeless ritual in which a "passive" subject has things happen to him but can contribute little of an instrumental nature himself, classical conditioning may be viewed fruitfully as an "active" or "dynamic" interaction between the organism and its environment. While the pejorative use of such terms as passive and "automatic" has little scientific significance, the possibility that the functionalistic approach described in this paper may lead us closer to the truth cannot be foreclosed merely because there is nothing "wrong" with the traditional view.

REFERENCES

BAXTER, R. Diminution and recovery of the UCR in delayed and trace classical GSR conditioning. *Journal of Experimental Psychology,* 1966, **71**, 447–451.

GRINGS, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychological Review,* 1960, **67**, 243–252.

GRINGS, W. W., LOCKHART, R. A., & DAMERON, L. E. Conditioning autonomic responses of mentally subnormal individuals. *Psychological Monographs,* 1963, **76**(39, Whole No. 558).

HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* (Rev. by G. A. Kimble) New York: Appleton-Century-Crofts, 1961.

KIMBLE, G. A., & DUFORT, R. H. The associative factor in eyelid conditioning. *Journal of Experimental Psychology,* 1956, **52**, 386–391.

KIMBLE, G. A., & OST, J. W. P. A conditioned inhibitory process in eyelid conditioning. *Journal of Experimental Psychology,* 1961, **61**, 150–156.

KIMMEL, H. D. Amount of conditioning and intensity of conditioned stimulus. *Journal of Experimental Psychology,* 1959, **58**, 283–288.

KIMMEL, H. D. The relationship between direction and amount of stimulus change and amount of perceptual disparity response. *Journal of Experimental Psychology,* 1960, **59**, 68–72.

KIMMEL, H. D. Further analysis of GSR conditioning: A reply to Stewart, Stern, Winokur, and Fredman. *Psychological Review,* 1964, **71**, 160–166.

KIMMEL, H. D. Instrumental inhibitory factors in classical conditioning. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium.* New York: Appleton-Century-Crofts, 1965. Pp. 148–171.

KIMMEL, H. D., & BAXTER, R. Avoidance conditioning of the GSR. *Journal of Experimental Psychology,* 1964, **68**, 482–485.

KIMMEL, H. D., & FOWLER, R. L. Disinhibition in compound conditioning as a function of the number of reinforcements of a single conditioned stimulus. Technical Report No. 3, 1961, University of Florida, Department of Psychology, Contract No. 580(09), Office of Naval Research.

KIMMEL, H. D., & GREENE, W. A. Disinhibition in GSR conditioning as a function of the number of CS–UCS trials and temporal location of the novel stimulus. *Journal of Experimental Psychology,* 1964, **68**, 567–572.

KIMMEL, H. D., & PENNYPACKER, H. S. Conditioned diminution of the unconditioned response as a function of the number of reinforcements. *Journal of Experimental Psychology,* 1962, **64**, 20–23.

KIMMEL, H. D., & SCHULTZ, C. A., JR. GSR magnitude and judgments of shock intensity as a function of physical intensity of shock. *Psychonomic Science,* 1964, **1**, 17–18.

KUPALOV, P. S. Some normal and pathological properties of nervous processes in the brain. *Annals of the New York Academy of Sciences,* 1961, **92**, 1046–1053.

LYKKEN, D. T. Preliminary observations concerning the "preception" phenomenon. American Psychological Association Division of Physiological and Comparative Psychology, *Newsletter,* **5**(2), 1959.

MCALLISTER, W. R. Eyelid conditioning as a function of the CS–UCS interval. *Journal of Experimental Psychology,* 1953, **45**, 417–422.

MERYMAN, J. J. *The magnitude of an unconditioned GSR as a function of fear conditioned at a long CS–UCS interval.* (Doctoral dissertation, State University of Iowa) Ann Arbor, Mich.: University Microfilms, 1953, No. A54–266.

MOELLER, G. The CS–UCS interval in GSR conditioning. *Journal of Experimental Psychology,* 1954, **48**, 162–166.

MORROW, M. C. Recovery of conditioned UCR diminution following extinction. *Journal of Experimental Psychology,* 1966, in press.

OST, J. W. P. *Interstimulus interval and proportion of non-reinforced trials in differential eyelid conditioning.* (Doctoral dissertation, Duke University) Ann Arbor, Mich.: University Microfilms, 1960, No. 60–6037.

PAVLOV, I. P. *Conditioned reflexes.* London: Oxford University Press, 1927.

PROKASY, W. F., FAWCETT, P. T., & HALL, J. F. Recruitment, latency, magnitude, and amplitude of the GSR as a function of interstimulus interval. *Journal of Experimental Psychology,* 1962, **64**, 513–518.

RODNICK, E. H. Characteristics of delayed and trace conditioned responses. *Journal of Experimental Psychology,* 1937, **20**, 409–425. (a)

RODNICK, E. H. Does the interval of delay of conditioned responses possess inhibitory properties? *Journal of Experimental Psychology,* 1937, **20**, 507–527. (b)

SMITH, K. Conditioning as an artifact. *Psychological Review,* 1954, **61**, 217–225.

SPENCE, K. W., & RUNQUIST, W. N. Temporal effects of conditioned fear on the eyelid reflex. *Journal of Experimental Psychology,* 1958, **55**, 613–616.

TOLMAN, E. C. A reply to Professor Guthrie. *Psychological Review,* 1938, **45**, 163–164.

WENGER, M. A. External inhibition and disinhibition produced by duplicate stimuli. *American Journal of Psychology,* 1936, **48**, 446–456.

WHITE, C. T., & SCHLOSBERG, H. Degree of conditioning of the GSR as a function of the period of delay. *Journal of Experimental Psychology,* 1952, **43**, 357–362.

(Received February 8, 1965)

Psychological Review
1966, Vol. 73, No. 3, 241

THEORETICAL NOTES

REPLY TO PAUL THOMAS YOUNG

FRANCIS W. IRWIN

University of Pennsylvania

Young (1966) claimed that there is evidence against Irwin's (1958) assertion that if an organism exhibits a discrimination it also exhibits a preference. The claim seems to be due to his not taking into account that, in any experimental determination, the objects of discrimination are not identical with the objects of preference.

In a recent paper Young (1966, p. 74) writes:

> If an animal shows a preference between two fluids, we can be sure that he is capable of discriminating between them, but the converse of this proposition is not necessarily true.

To this he appends the following footnote:

> Irwin (1958) wrote: "By our definition, an organism cannot exhibit a preference without exhibiting a discrimination, and conversely [p. 162]." The first part of this statement is certainly true, but the converse is false. There is plenty of evidence that animals and human beings discriminate without revealing a preference.

This flat denial of a main point of my paper (Irwin, 1958) by a well-known student of preferences requires an answer. The point itself is not trivial, even if its implications for method are put aside. Discrimination has long been recognized as a fundamental psychological concept; my conclusion, based in part upon the assertion that Young denies, was that the motivational concept of preference is equally fundamental.

If my statement meant what Young seems to have taken it to mean, I would indeed be convicted, not of going against the facts—the question is one of choice of defining criteria, not one of fact—but of adopting criteria that were, to say the least, esoteric. I agree with Young's implication that an animal may discriminate between two fluids without preferring one to the other, but his quotation from my paper does not contradict this. What the whole argument of my paper leads to is the implication that if an animal exhibits a discrimination between two fluids, then, because of the relations between the criteria offered for discrimination and preference, it also exhibits a preference for *some outcome over some other outcome*. Indeed, the next two sentences of my article read:

> It is important to recognize, however, that in a given experimental determination, the objects of preference are never identical with the objects of discrimination. The former are features of the outcomes of acts, while the latter are features of the situations that exist before the organism acts [p. 162].

I regret Young's misunderstanding of my paper all the more because I believe that he and I are allied closely in our views of motivational problems. Were he to perform an experiment to determine whether animals would discriminate between some pair of discriminanda, I am confident that he would provide differential outcomes ("incentives")—perhaps a highly palatable sucrose solution and an empty vial—that he was already sure would be highly different in preference value for the animals. If they displayed the discrimination, they would necessarily also show that his judgment of the preference was correct.

REFERENCES

IRWIN, F. W. An analysis of the concepts of *discrimination* and *preference*. *American Journal of Psychology,* 1958, **71**, 152–163.

YOUNG, P. T. Hedonic organization and the regulation of behavior. *Psychological Review,* 1966, **73**, 59–86.

(Early publication received February 7, 1966)

Psychological Review
1966, Vol. 73, No. 3, 242–247

ON NOISE IN THE NERVOUS SYSTEM[1]

LAWRENCE R. PINNEO

Delta Regional Primate Research Center, Tulane University

Treisman and Hebb have suggested that "spontaneous," "random," or "background" activity in the nervous system constitutes "noise" in discrimination and learning; that is, this type of activity has no functional value to the organism. This paper attempts to show that tonic activity, a term including all of the types of activity listed above, is rather the functional substrate of the brain. Examples are cited for the skeletal and autonomic motor systems, the primary sensory systems, and the diffuse ascending and descending reticular activating systems to show that the tonic activity in the entire brain enters into all discrimination and learning, and, in agreement with Lashley, represents the neural basis of behavior.

Two recent theoretical papers have argued that "noise" in the nervous system has interfering effects in discrimination and learning. In one, Treisman (1964) has suggested that three sources of noise limit discrimination: (*a*) the irreducible physical variability of the stimulus, (*b*) the "spontaneous" neural background activity to which a stimulus is added, and (*c*) the neural noise arising from variation in the pathways transmitting messages centrally. Based on these three sources of noise, Treisman has worked out a complicated signal-detection theory to explain the form of the Weber function for visual intensity discrimination and for other sense dimensions.

In the other paper, Hebb (1961) raised the question of the interfering effects of random activity in the nervous system during learning of a specific task. By learning he meant the modification of the direction of transmission in the central nervous system (CNS) at the synapse. He pointed out that a large brain such as a mammal's has many more neurons present than are necessary for learning a specific task. Therefore, random activity in the neurons not involved in learning the task constitutes noise, which Hebb felt must interfere with learning.

These two papers illustrate a widely held misconception of brain function, namely, that the spontaneous, random, or background discharge of neurons has little or no functional value; that is, this activity has no information value for the organism and therefore is noise in the communications sense of the word. In this theoretical note I attempt to show that the *neural* noise to which Treisman and Hebb refer (leaving out Treisman's first category) is not noise at all, that this neural activity does not limit discrimination nor interfere with learning, and that in fact this activity is essential to discrimination and learning.

Arduini (1963) has suggested that there are fundamentally two types of nervous discharge, and he has carefully defined them; his definitions will be used in this paper. Borrowing from the terminology applied to muscle activity, Arduini defines "phasic" activity as a transient increase or decrease in impulse firing rates of neurons that is time locked to a particular stimulus. A familiar example is the evoked response. "Tonic" activity on the other hand is nontransient, or continuous, neural discharge in which the average firing rate is random and constant and is *not* time locked to a stimulus. So-called spontaneous activity (a dubious term at best since it implies the discharge of neurons without benefit of influences external to the neurons), background activity, random discharge, and the like, are familiar examples of tonic activity.

---

[1] This work was supported by National Institutes of Health, United States Public Health Service Grants FR-00164-03 and NB-04951-01.

Tonic activity may conveniently be considered as the functional substrate of the nervous system. As Lashley (1951) has pointed out, neurological theory has been dominated by the incorrect belief that neurons of the CNS are inactive for the greater part of the time. According to this view, neurons are always in the resting state where they are linked in the relative isolation of reflex arcs until a particular reaction for which they are specifically associated activates them. Yet, no matter where in the CNS one places an electrode, and no matter what the condition of the organism as long as he is alive, a steady discharge of neurons may always be recorded. In general, level of "arousal" or "activation" increases with the level of tonic activity throughout the brain. And it is against this tonic level that sensory messages are received and motor acts achieved. As Lashley (1951, p. 131) has so well expressed it, "all the cells of the cerebrospinal axis are being continually bombarded by nerve impulses from various sources and are firing regularly, probably even during sleep." And, "in the intact organism, behavior is the result of interaction of this background of excitation with input from any designated stimulus [p. 112]."

The nonspecific attributes of tonic activity are now well documented following the experiments of Moruzzi and Magoun (1949) on the cortical arousing effects of stimulation of the brain-stem reticular formation, and the descending tonic influence of the lower brain stem upon postural tonus as shown by Lindsley, Shreiner, and Magoun (1949). Even tonic motor activity itself increases level of activation and modifies behavior (Pinneo, 1961). But tonic activity also has been shown to have specific functional value in the skeletal and autonomic motor systems for the regulation of movement and the function of visceral organs. Thus, Sherrington (1906) observed that at the level of motor units the intensity of a mechanical effect resulting from a normal tonic discharge depends in large part upon the frequency of tonic impulses. Tower (1949) has carried this concept a little further by suggesting that the tonic nature of the discharging motor neurons is the expression of a central excitatory state (cf. Sherrington, 1906), modulated by a phasic or episodic function which appears as a specific contribution to individual acts.

Stimulation of central tracts and nuclei concerned with regulation of the skeletal and autonomic motor systems illustrates the *fine* control that tonic activity in these systems exerts (see the many chapters devoted to these systems in Field, Magoun, & Hall, 1960). Almost countless locations have been found (for example, in the pyramidal and extrapyramidal systems, cerebellum, and brain stem) in which stimulation with gross electrodes over short and long periods of time appears to initiate and maintain highly integrated and complex patterns of muscle activity. Frequently, it is found that the extent and complexity of movement is directly related to current strength. What moves, however, depends upon the location of the electrode. Since at these levels it is highly unlikely that fibers are excited selectively and in the proper temporal sequence to produce such complex and organized movements, it must be concluded that it is the increase in tonic level brought about by stimulation of the descending tracts that controls the motor activity. The argument (Patton & Amassian, 1960) that many of these movements are reflex results of afferent stimuli by spread of current from efferent tract stimulation to nearby afferent tracts has some validity. However, stimulation of efferent pathways alone can produce the same highly organized movements (Dow & Moruzzi, 1958; Pompeiano, 1959).

In our laboratory we have repeated and extended some of these experiments and have been able to completely reproduce a series of highly integrated movements in the squirrel monkey (*Saimiri sciureus*) by simultaneous and successive stimulation of several critical nuclei in the cerebellum and brain stem (Pinneo, in press—a). First, with a single stimulating electrode pair, we explored various loci in such nuclei as the interpositus, Deiter's, fastigial, and vestibular, and several other nuclei and tracts of the caudal brain stem, to determine the type of movement that could be elicited at each locus and the optimum electrical parameters for obtaining the movement. Then, by chronically implanting up to six pairs of electrodes in chosen sites, and by the proper ordering of stimulation at each electrode pair with respect to the other electrodes, fine coordinated movements were reproduced at our desire.

These effects were obtained whether the motor cortex concerned with the muscular groups in question was intact or not. For example, destruction of the motor cortex for the right arm and hand of the squirrel monkey produces paralysis in that limb, as in man. But, by proper placement and stimulation of electrodes in the homolateral interpositus nucleus, we were able to close the fingers (as in a grasp around an object) and to bring the hand rapidly and smoothly to the mouth; this could be repeated as often as desired. Other possible combinations included extension of the arm outward (as in reaching), backward, and to the side. Other electrode locations were found that produced similar movements in other limbs, opening and closing of the mouth (as in mastication), movement out and in of the tongue (as in licking), and control of rate and amount of respiration, and rate and volume of cardiac output. In all of the locations examined, the mechanical effect was proportional to current strength in the stimulating electrode (and therefore proportional to the level of tonic activity in the tracts or nuclei involved), so that very finely graded movements were obtainable singly or in combination.

The functional value of tonic activity is not limited to the skeletal and autonomic motor systems. In the primary visual pathway of mammals, tonic activity alone is recorded under maintained states of darkness or light (Arduini & Pinneo, 1961, 1962, 1963b; Granit 1955). In the dark-adapted state a large number of retinal ganglion cells, perhaps most, maintain a relatively high rate of tonic discharge (Arduini & Cavaggioni, 1961; Granit, 1955). With the onset and offset of a prolonged light stimulus, "on" and "off" retinal bursts provide a momentary phasic response throughout the visual system whose size is proportional to the stimulus intensity. Shortly thereafter, however, and for as long as the intensity of the stimulus remains constant, the activity in the primary visual system is tonic (Arduini & Pinneo, 1961, 1962, 1963b). Furthermore, the level of the tonic activity with a steady light was found by Arduini and Pinneo (1962) to be inversely related to the intensity of the light by a power function of a form quite similar to the brightness function described by Stevens (1962) for humans.

Recently, Brooks (1964) has confirmed that the level of tonic activity in the primary visual system is the probable neural basis of brightness vision. Using the classical $\Delta I/I$ versus I paradigm, Brooks trained squirrel monkeys to make brightness discriminations over 4 log units of background light and then recorded the phasic and tonic activity (due to ΔI and I, respectively) from the lateral geniculate nucleus under threshold conditions. Brooks found that (*a*) the Weber function of the squirrel monkey was very similar to that of man; (*b*) the tonic and phasic activity changed with illumination of the retina as Arduini and Pinneo (1962, 1963b) had shown for the cat; and (*c*) brightness discrimination at threshold was inversely proportional to the level of tonic activity in the geniculate. Thus, in primates as in the cat, the noise in the visual system to which Treisman (1964) refers is *highest* during dark adaptation when visual sensitivity is greatest and *lowest* under conditions of high illumination when sensitivity is poorest. It is difficult, therefore, to see how this noise interferes with discrimination.

Perception of flicker fusion and the brightness of an intermittent light at critical fusion frequency (CFF—the Talbot brightness) also have been shown to depend upon the presence and level of tonic activity in the visual system (Arduini & Pinneo, 1963a). Recording from electrodes implanted in human patients for therapeutic reasons, Pinneo (in press—b) and Heath found that CFF, as determined by verbal report, occurred at that frequency of intermittent light where all or most phasic activity in the visual system had disappeared and only tonic activity remained. Furthermore, they found that the level of tonic activity at CFF was the same as that produced by a steady light of the same intensity as the average intensity of the flickering light. Thus, the flicker was perceived as steady because the activity of the visual system responded as though the light were steady; that is, the discharge was tonic, and the level of tonic activity was inversely proportional to the power of the average intensity.

Functional relations also have been found between the level of tonic activity in other primary sensory systems and environmental stimuli, notably the auditory pathway (Galambos & Davis, 1944; Galin, 1964) and the temperature organ (Hensel, 1952; Zotterman, 1953). Though many other examples could be cited, these and the others described above should suffice to illustrate that tonic activity in the nervous system has very specific functional value and therefore is not merely noise.

Analysis of the mode of action of phasic and tonic activity in the CNS as demonstrated by the experiments discussed suggests that phasic activity selects and guides responses while tonic activity controls the fineness and extent of the response. In a given sequence from stimulus to response, the distribution of the two types of activity throughout the CNS may be described as follows:

1. In sensory systems phasic activity is the response to transient stimuli, while the level of tonic activity signals the average intensity of background stimuli.

2. Both phasic and tonic activity in sensory systems modulate the existing tonic discharge in the brain-stem reticular formation. In this case the state of arousal is modified by the level of tonic activity in the ascending reticular activating system; postural tonus is changed by the tonic discharge of the descending reticular activating system (motor set to respond?), while the phasic discharges probably mediate attention or orienting responses.

3. In the skeletal and autonomic motor systems the phasic activity in specific motor nuclei and tracts selects and initiates movement and modifies visceral function while tonic activity provides the fine control and extent of movements and level of visceral function.

4. Feedback of phasic and tonic activity in all systems—receptor, sensory, motor, autonomic, and activating—to and from the cerebral cortex and subcortical nuclei provides the continued sequence of responses necessary for completion of an act.

It is evident from this analysis that stimuli are always superimposed upon the tonic activity of the *entire* brain, and that the perception and "meaning" of a stimulus must depend upon the relative amounts of tonic activity in the various parts of the nervous system. Discrimination, generalization, and learning occur with repeated exposures under different and similar circumstances with different and similar patterns of tonic activity throughout the nervous system. This is well illustrated by the fact that perception, discrimination, and learning depend upon and may be modified by changes in level of activation, patterns of muscle activity, and states of visceral activity and the activity in other sensory systems, all primarily tonic in nature.

In his thoughtful evaluation of cerebral organization and behavior, Lashley (1958) has given to tonic activity the major role of the neural basis of mind. Though reminiscent of Hebb's (1949) concept of the "cell assembly" and "phase sequence," in that Lashley organizes the brain into a large number of systems where each system consists of the traces of a number of habits or memories, Lashley still ascribes mental function to the tonic activity of these trace systems rather than to a particular set of connections between specific neurons and their patterns of activity. This concept is not only consistent with what is now known of the physiology of the brain, but is also the only explanation that will satisfy the efficiency of behavior following brain damage (e.g., Lashley's laws of "mass action" and "equipotentiality") and the many controversial neurological interpretations of events in perception and learning.

REFERENCES

ARDUINI, A. The tonic discharge of the retina and its central effects. In G. Moruzzi, A. Fessard, & H. H. Jasper (Eds.), *Progress in brain research.* Vol. 1. *Brain mechanisms.* New York: Elsevier, 1963. Pp. 184–206.

ARDUINI, A., & CAVAGGIONI, A. Attività tonica della retina registrata con semimicroelettrodi. *Bollettino della Società italiana di biologia sperimentale,* 1961, **37,** 1393–1395.

ARDUINI, A., & PINNEO, L. R. Attività nel nervo ottico e nel genicolato laterale nell'oscurita 'e durante l'illuminazione continua. *Bollettino della Società italiana di biologia sperimentale,* 1961, **37,** 430–432.

ARDUINI, A., & PINNEO, L. R. Properties of the retina in response to steady illumination. *Archives italiennes de biologie,* 1962, **100,** 425–448.

ARDUINI, A., & PINNEO, L. R. The effects of flicker and steady illumination on the activity of the cat visual system. *Archives italiennes de biologie,* 1963, **101,** 508–529. (a)

ARDUINI, A., & PINNEO, L. R. The tonic activity of the lateral geniculate nucleus in dark and light adaptation. *Archives italiennes de biologie,* 1963, **101,** 493–507. (b)

BROOKS, B. A. *Neural correlates of brightness discrimination in the squirrel monkey* (Saimiri sciureus). (Doctoral dissertation, Florida State University) Ann Arbor, Mich.: University Microfilms, 1964, No. 65-326.

DOW, R. S., & MORUZZI, G. *The physiology and pathology of the cerebellum.* Minneapolis: University of Minnesota Press, 1958.

FIELD, J., MAGOUN, H. W., & HALL, V. E. (Eds.), *Handbook of physiology: Section I. Neurophysiology.* Vol. 2. Washington, D.C.: American Physiological Society, 1960.

GALAMBOS, R., & DAVIS, H. Inhibition of activity in single auditory nerve fibers by acoustic stimulation. *Journal of Neurophysiology,* 1944, **7,** 283–303.

GALIN, D. Auditory nuclei: Distinctive response patterns to white noise and tones in unanesthetized cats. *Science,* 1964, **146,** 270–272.

GRANIT, R. *Receptors and sensory perception.* New Haven: Yale University Press, 1955.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HEBB, D. O. Distinctive features of learning in the higher animal. In A. Fessard, R. W. Gerard, J. Konorski, & J. F. Delafresnaye (Eds.), *Brain mechanisms and learning.* Springfield, Ill.: Charles C Thomas, 1961. Pp. 37–51.

HENSEL, H. Physiologie der Thermoreception. *Ergebnisse der Physiologie (biologischen Chemie und experimentellen Pharmakologie),* 1952, **47,** 165–368.

LASHLEY, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley, 1951. Pp. 112–147.

LASHLEY, K. S. Cerebral organization and behavior. *Proceedings of the Association for Research in Nervous and Mental Diseases,* 1958, **36,** 1–18.

LINDSLEY, D. D., SCHREINER, L. H., & MAGOUN, H. W. An electromyographic study of spasticity. *Journal of Neurophysiology,* 1949, **12,** 197–205.

MORUZZI, G., & MAGOUN, H. W. Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology,* 1964, **1,** 455–473.

PATTON, H. D., & AMASSIAN, V. E. The pyramidal tract: Its excitation and functions. In J. Field, H. W. Magoun, & V. E. Hall (Eds.), *Handbook of physiology: Section I. Neurophysiology.* Vol. 2. Washington, D.C.: American Physiological Society, 1960. Pp. 837–861.

PINNEO, L. R. The effects of induced muscle tension during tracking on level of activation and on performance. *Journal of Experimental Psychology,* 1961, **62,** 523–531.

PINNEO, L. R. Electrical control of behavior by programmed stimulation of the brain. *Nature,* in press. (a)

PINNEO, L. R. Neurology of brightness vision. *Scientific American,* in press. (b)

POMPEIANO, O. Organizzazione somatotopica delle risposte flessorie alla stimolazione elettrico del nucleo interposito nel gatto decerebrato. *Archivio di scienze biologiche,* 1959, **43**, 163–176.

SHERRINGTON, C. S. *The integrative action of the nervous system.* New Haven: Yale University Press, 1906.

STEVENS, S. S. The surprising simplicity of sensory metrics. *American Psychologist,* 1962, **17**, 29–39.

TOWER, S. S. The pyramidal tract. In P. C. Bucy (Ed.), *The precentral motor cortex.* Urbana: University of Illinois Press, 1949. Pp. 149–172.

TREISMAN, M. Noise and Weber's law: The discrimination of brightness and other dimensions. *Psychological Review,* 1964, **71**, 314–330.

ZOTTERMAN, Y. Special senses: Thermal receptors. *Annual Review of Physiology,* 1953, **15**, 357–372.

(Received January 4, 1965)

Psychological Review
1966, Vol. 73, No. 3, 248–251

ON REMOTE ASSOCIATIONS AND THE INTERPRETATION OF DERIVED–LIST EXPERIMENTS[1]

DAVID T. HAKES AND ROBERT K. YOUNG

University of Texas

Serial learning is generally assumed to involve the formation of direct and remote interitem associations. Such an interpretation appears to predict positive transfer in the derived-list paradigm, a prediction apparently supported by Ebbinghaus' finding of savings in this paradigm. It is pointed out that: (a) The theory actually predicts negative transfer in this paradigm due to interference from associations which were correct in the 1st list and are incorrect in the 2nd list; and (b) because of the control condition used, Ebbinghaus' findings indicate negative or 0 transfer. Recent findings of negative or 0 transfer are consistent with the theory and with the Ebbinghaus data when both are correctly interpreted.

Since Ebbinghaus' (1885) first experiments, the dominant conception of serial learning has been that it involves the formation of interitem associations between the adjacent items of the list. Thus the learning of a serial list, A-B-C- . . . -*N*, is hypothesized to involve the formation of associations A-B, B-C, etc. In addition, serial learning theorists have generally assumed that remote as well as direct interitem associations are formed (cf. McGeoch & Irion, 1952, Ch. 4). Formation of remote associations is assumed to occur in both the forward (e.g., A-C, A-D) and backward (e.g., B-A, C-A) directions, and their strength is assumed to decrease as the distance between the items increases, forward associations being stronger than backward associations of the same degree of remoteness.

Through these assumptions concerning remote associations, it is possible to derive an account of the bowed serial-position curve typically found during serial learning (Bugelski, 1950). The greater difficulty of items in the middle of the list is attributed to the greater number and strength of remote associations for these items than for those at the ends of the list, and the asymmetry of the serial-position curve is attributed to the greater strength of (and, hence, greater interference from) forward than backward associations. While some theorists have suggested modifications of such a theory (e.g., Ribback & Underwood, 1950), few have been led to question the assumption of remote associations. Acceptance has been so widespread over such a period of time that Slamecka (1964) has suggested that it might be referred to as the "doctrine" of remote associations.

A major early source of support for the remote-association assumption was the results of Ebbinghaus' (1885) studies of the derived-list paradigm. Ebbinghaus found savings in second, derived-list learning which were greatest for the first-order condition and decreased as the order of derivation (i.e., degree of remoteness) increased. The smallest amount of savings was obtained when the derived list was a random rearrangement of the items. These results have generally been interpreted as demonstrating the existence of remote forward associations and as indicating that these decrease in strength as the degree of remoteness increases.

In view of the methodological shortcomings of Ebbinghaus' experiments (cf. Hakes, James, & Young, 1964; Slamecka, 1964), it is somewhat surprising that their results and interpretation have been so widely accepted without serious question or adequate replication. It may well be that their acceptance resulted from their apparent consistency with prevailing serial learning theory. Three early studies of the derived-list paradigm (Cason, 1926; Hall, 1928; Lepley, 1934) yielded results which were not very supportive of Ebbinghaus' data. But because of methodological problems, these studies have been accorded little weight in the literature (Slamecka, 1964, p. 63).

---

[1] Preparation of this note was supported in part by National Science Foundation Grant GB-979.

Irion (1946) did report that the random-order derived-list condition (the one used by Ebbinghaus as a control condition) yielded negative transfer and retroactive inhibition. In commenting upon this finding, McGeoch and Irion (1952) note that:

> This interference cannot serve except to reduce remote association effects. The positive evidence for remote associations is not a function of such interference, although remote associations, themselves, may be a function of *positive* transfer of training [p. 94].

This would appear to represent an attempt to reconcile Ebbinghaus' finding of savings with Irion's finding of interference, but something seems to be lacking.

Recently, a number of experiments have been performed on the derived-list paradigm using more adequate methodology than the early studies. Hakes et al. (1964), studying first-, second-, and third-order derived-list conditions, obtained negative transfer which was greatest for the first-order condition and decreased as the order of derivation increased. Young, Hakes, and Hicks (1965), studying 8-, 12-, and 16-item lists in the first-order derived-list condition, found strong evidence of negative transfer for the 12- and 16-item conditions and nonsignificant positive transfer for the 8-item condition.

Slamecka (1964) has performed a series of derived-list experiments involving the first-order condition and a condition involving, within a single list, zero-, first-, and second-order conditions. Comparing these conditions with a random-order condition, Slamecka reported no evidence of faster learning of the derived lists except in one case, involving the first-order condition, in which the subjects were explicitly told the relationship between first and second lists.

It is not the question of the empirical results of derived-list studies which is of interest here. This question would appear to be answered adequately by the experiments just discussed. The question is, rather, how these results are to be interpreted. With reference to this, two questions arise. The first is concerned with the relation between Ebbinghaus' results and those of more recent experiments. The immediately apparent conclusion is that the finding of negative transfer is incompatible with Ebbinghaus' results. As Hakes et al. (1964) state: "It would appear, then, that the present experiment provides no support for Ebbinghaus' data or for the theory of remote associations based on these data [p. 513]." On closer inspection, however, such a conclusion appears not to be justified.

It has generally been assumed that Ebbinghaus' results, reported in terms of savings, showed positive transfer (cf. McGeoch & Irion, 1952, p. 94). Since Irion (1946) has shown that the random-order condition (Ebbinghaus' control condition) yields *negative transfer,* Ebbinghaus' findings cannot be considered to show positive transfer: The savings in the derived-list conditions were greater than for a negative-transfer condition and might well represent decreased negative transfer rather than positive transfer. Thus it would appear that the inconsistency between the savings found by Ebbinghaus and the negative transfer found by Hakes et al. (1964) and by Young et al. (1965) is more apparent than real.

Slamecka's results may be interpreted in the same manner. His general finding is that the derived-list condition is not superior to the random-order condition and is, thus, a negative-transfer condition. It is of interest to note that when subjects are aware of the interlist relationship in the first-order derived-list paradigm the amount of negative transfer is reduced. This condition would appear to be directly comparable to the Ebbinghaus experiments, as Ebbinghaus used himself as the only subject and must have been aware of the relationship between lists. The preponderance of evidence suggests, then, that the derived-list paradigm yields negative transfer. Instruction can reduce the amount of transfer, but whether under these conditions positive transfer occurs is indeterminate since neither Ebbinghaus nor Slamecka used appropriate control conditions.

The second, and major, issue of interpretation concerns the relationship of these results to an interitem theory of serial learning. Initially, it would seem that the results are inconsistent with such a theory, at least with one which assumes remote associations. Again, closer inspection reveals that this is not the case.

It has been assumed that the formation of remote associations during first-list learning should facilitate the learning of a second list in which the remote associations become the direct interitem associations. That is, during second-list learning, the previously incorrect, remote associations become direct, correct associations, and the strength which they have accrued during first-list learning is the basis for positive transfer. Such an interpretation of the remote-association hypothesis is, of course, consistent with the erroneous interpretation of Ebbinghaus' results as indicating positive transfer.

Such an interpretation of the hypothesis is, however, incomplete. In the learning of a derived second list, the previously correct, direct associations are no longer correct and are instead remote associations. These associations must, at the end of first-list learning, have been stronger than any remote associations in order for the first list to have been learned. During second-list learning, these strong, and now incorrect, associations should, according to an interitem-association theory, be a source of powerful interference and negative transfer. Thus it would appear that a theory of direct and remote interitem associations considers the derived-list paradigm to be a negative-transfer paradigm. Positive transfer can be predicted only if interference from correct first-list associations is ignored. It should be noted that when an analogous paired-associate learning situation is encountered (Young & Jennings, 1964), in which associations incorrect in the first list become correct in the second, negative transfer is obtained.

It would appear, then, that contrary to the customary interpretation, the remote-association hypothesis is consistent with the obtained negative transfer in the derived-list paradigm. The confusion which has arisen has two sources: the misinterpretation of the Ebbinghaus results and an improper derivation from the hypothesis. When both errors are made, Ebbinghaus' results appear to be consistent with the hypothesis. The inconsistency becomes apparent only upon consideration of the recent experiments which measured transfer rather than savings.

The one inconsistency which remains is that the remote-association hypothesis appears to predict that the amount of negative transfer should increase as the degree of remoteness increases. This, of course, is the result which Ebbinghaus obtained. The one recent study which yields data relevant to this prediction (Hakes et al., 1964) does not support the prediction. No explanation for this is readily available.

Finally, it should be noted that the empirical results which have been discussed are also consistent with other interpretations of serial learning, such as an ordinal position theory (cf. Young, 1962).

REFERENCES

BUGELSKI, B. R. A remote association interpretation of the relative difficulty of learning nonsense syllables in a serial list. *Journal of Experimental Psychology,* 1950, **40,** 336–348.

CASON, H. Specific serial learning; a study of remote forward associations. *Journal of Experimental Psychology,* 1926, **9,** 299–324.

EBBINGHAUS, H. *Uber das Gedächtnis: Untersuchungen zur experimentellen Psy-*

chologie. Leipzig: Duncker & Humblot, 1885.

HAKES, D. T., JAMES, C. T., & YOUNG, R. K. A re-examination of the Ebbinghaus derived-list paradigm. *Journal of Experimental Psychology,* 1964, **68,** 508–514.

HALL, M. E. Remote associative tendencies in serial learning. *Journal of Experimental Psychology,* 1928, **11,** 65–76.

IRION, A. L. Retroactive inhibition as a function of the relative serial positions of the original and interpolated items. *Journal of Experimental Psychology,* 1946, **36,** 262–270.

LEPLEY, W. M. Serial reactions considered as conditioned reactions. *Psychological Monographs,* 1934, **46**(1, Whole No. 205).

MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

RIBBACK, A., & UNDERWOOD, B. J. An empirical explanation of the skewness of the bowed serial position curve. *Journal of Experimental Psychology,* 1950, **40,** 329–335.

SLAMECKA, N. J. An inquiry into the doctrine of remote associations. *Psychological Review,* 1964, **71,** 61–76.

YOUNG, R. K. Tests of three hypotheses about the effective stimulus in serial learning. *Journal of Experimental Psychology,* 1962, **63,** 307–313.

YOUNG, R. K., HAKES, D. T., & HICKS, R. Y. Effects of list length in the Ebbinghaus derived-list paradigm. *Journal of Experimental Psychology,* 1965, **70,** 338–341.

YOUNG, R. K., & JENNINGS, P. C. Backward learning when the same items serve as stimuli and responses. *Journal of Experimental Psychology,* 1964, **68,** 64–70.

(Received January 19, 1965)

Psychological Review
1966, Vol. 73, No. 3, 252–256

A COMMENT ON INTERMEDIATE SIZE DISCRIMINATION AND ADAPTATION-LEVEL THEORY[1]

DONALD A. RILEY, MARIAN SHERMAN,[2] AND JOHN P. McKEE

University of California, Berkeley

Zeiler's (1963a, 1963b) adaptation-level (AL) interpretation of transposition in children following training on an intermediate size problem is compared with Stevenson and Bitterman's (1955) conclusion that S learns to respond to both relational and absolute properties of the stimulus. Based both on the analysis of Zeiler's data and new data it is concluded that while it is not necessary to assume a changing to AL to account for the facts of intermediate size learning and transfer, it is necessary to assume that S's response in training and test is determined both by the absolute and relative properties of the stimuli. The degree to which these properties control behavior depends on factors such as background characteristics and instructions.

The purpose of this note is threefold. First Zeiler's (1963b) account of transposition after intermediate size training in terms of adaptation level (AL) will be compared with an interpretation suggested by Stevenson and Bitterman (1955). Second, the authors will comment briefly on Zeiler's data and their relevance to both positions. Finally some further research on the intermediate size problem which was designed to replicate and extend Zeiler's work will be presented.

Zeiler's (1963b) extension of AL theory differs from James' (1953) similar extension mainly in his assumption that the cue to which the subject (*S*) learns to respond during training is the ratio of the value of the correct training stimulus to the training AL. According to this hypothesis, transposition occurs because as soon as *S* sees the test stimuli his AL changes, and he then chooses that stimulus whose ratio to the *new* AL is nearest the ratio of the original positive stimulus to the *training* AL. By the assignment of appropriate weights to the training stimuli and the test stimuli, Zeiler is able to account for a variety of results from previous intermediate size experiments. In addition to rationalizing previous work, Zeiler (1963a, 1963b) has presented a large amount of new information which he argues is consistent with the AL interpretation, but not with previous analyses of transposition.

A COMPARISON OF TWO INTERPRETATIONS

Stevenson and Bitterman (1955) interpreted the results of their experiment on transposition following intermediate size training as evidence that *S*s learned to respond to both relational and absolute properties of the stimuli. Although *S*s showed relational learning by transposing in the near test, the observed decline in transposition in a far test was interpreted as the consequence of the large difference between the absolute values of the stimuli in training and testing. Since Stevenson and Bitterman's report, a number of other papers in the last few years have identified situational factors that influence the degree to which *S*s respond to relational or absolute properties of the stimuli. For instance, Baker and Lawrence (1951) have shown the importance of temporal factors, and Riley,

---

[1] This research was supported in part by Grant HD00360-02 from the United States Public Health Service and in part by a grant from the National Science Foundation to the Institute of Human Learning.

We are indebted to Maxine Christopher of the Oakland Public Schools for making subjects available to us. We sincerely appreciate the friendly help and cooperation of the directors and teachers of the Child Care Centers.

[2] United States Public Health Service (NIMH) Predoctoral Fellow.

Goggin, and Wright (1963) have shown the importance of variations of the spatial arrangements of the stimuli.

Stevenson and Bitterman's position accounts for the amount and direction of transfer in terms of (*a*) what is associated during training, and (*b*) the similarity of training and test situations. This position differs from the AL account primarily in one way. It does *not* assume that transposition (or any other outcome of a transfer test) occurs because the condition of *S* changes while he is looking at the stimuli prior to making a choice. The AL account of transposition, as we understand it, attributes transposition in the intermediate size problem to a change in the perceived size of the stimuli. Stevenson and Bitterman's position denies neither the existence of such sensory-adaptation processes nor that under very special circumstances such shifts in the state of *S* might produce transposition. It does assert that its two assumptions are both necessary and sufficient for dealing with the facts revealed by transfer tests after training on the intermediate size problem and that the additional assumption of a change in adaptation level is superfluous.

ZEILER'S DATA

Zeiler's (1963b) article considers two kinds of evidence. One is the evidence from previous intermediate size experiments. Zeiler suggests an AL interpretation that is based on an ad-hoc weighting of the stimuli presented during training and testing. Such ad-hoc capabilities are, of course, only minimally convincing evidence for the validity of a theory.

More interesting are the results from Zeiler's own intermediate size experiments. His stimuli were flat, ¼-inch Masonite squares with indentations on the underside to hide the reward object. The smallest, Stimulus 1, was 4 square inches, and the squares increased in area by a factor of 1.4 to the largest, Stimulus 8, which was 42.2 square inches. These stimuli were presented on a 20 × 30-inch tray of heavy cardboard. A 16 × 20.5-inch cardboard screen was used to hide the stimuli between trials. In his first set of experiments, Zeiler reports two sorts of data: (*a*) those in which the training and test stimuli overlap at two points (e.g., 1-2-3 : 2-3-4), and (*b*) those in which the stimuli overlap at one point (1-2-3 : 3-4-5) or not at all (1-2-3 : 4-5-6). In the former experiments *S*s transpose when tested, while in the latter they choose the stimulus which is closest in value to the originally correct. For these experiments Zeiler is able to predict the test results by using the same weights that he had previously used to account for Stevenson and Bitterman's (1955) data.[3] Though such an achievement seems impressive, Zeiler's experiments are sufficiently similar to Stevenson and Bitterman's original ones that Stevenson and Bitterman's interpretation in terms of relative and absolute aspects of the stimulus situation remains equally tenable.

In his next set of 17 experiments Zeiler increases the size ratio, but the general nature of the problems remains the same. Only 7 of these experiments with large ratios between stimuli (e.g., 1-3-5 : 3-5-7) involve a prediction based on AL theory which is pitted against the prediction that *S* will choose the stimulus which was correct during training. In 6 of the 7, the AL predictions—based on the previous weighting—fail, and only 1 is confirmed. In all 6 failures *S*s select the originally correct stimulus.

Zeiler suggests that in determining AL the training stimuli should perhaps be weighted more heavily when they are very discriminable from each other, but he gives no rationale. Furthermore, in some of these cases the discriminability of the *test* stimuli is also much greater than in his early experiments. In any event, the simplest explanation of these data is that the response is under the control of the positive training stimulus and *not* of a shifting AL.

Zeiler (1963a) reports another sort of evidence which he interprets in AL terms. *S*s were trained with an asymmetrical array of stimuli and tested with an overlapping symmetrical array (e.g., 1-2-8 : 1-2-3). The training stimuli were always

---

[3] A wide range of weights will do equally well.

1-2-8, and the test stimuli for different groups were 1-2-6, 1-2-5, 1-2-4, and 1-2-3. *S*s tested with 1-2-6 overwhelmingly chose the originally correct Stimulus 2, but as the largest test stimulus decreased in size there was an increasing tendency to select Stimulus 1. Although this last result is in accord with the AL hypothesis, the results for all four conditions considered together are not. Zeiler suggests that perhaps asymmetrical conditions require different weights, but again the rationale is not clear.

Although the shift to Stimulus 1 is not predictable from Stevenson and Bitterman's (1955) hypothesis, an ex-post-facto interpretation is obvious. The presentation of a test set, 1-2-3, after training on 1-2-8 leaves one of the relationships undisturbed (i.e., 1-2), while another one (2-3 as opposed to 2-8) is radically changed. If during training *S* has learned, among other things, to choose a stimulus which is a particular amount smaller than the largest, then the substitution of 3 for 8 should force *S*'s choice toward 1. In short, these new phenomena are of real interest, but there is nothing in them that contradicts the assumption that *S* responds to both the absolute and the relational properties of the stimuli with the relative importance of these two factors determined by a variety of conditions during training and testing. But the fact remains that the shift from the intermediate stimulus to the smallest as the value of the largest stimulus declines is a dramatic effect predicted by the AL position and not merely rationalized after the fact.

TABLE 1

NUMBER OF SUBJECTS CHOOSING SMALL (1), INTERMEDIATE (2), AND LARGE (3) TEST STIMULUS AFTER TRAINING WITH STIMULI 1, 2, AND 8

Condition	Test choice			
	N	S	I	L
Zeiler's experiment	30	16	10	4
Replication of Zeiler's experiment	28	0	17	11
Instructions to attend to all 3 stimuli	13	0	9	4
Nonrandom training arrays	11	2	6	3

SOME NEW EMPIRICAL DATA

In view of the importance of this switch in preference the authors decided to repeat the most extreme condition of Zeiler's experiment—that is, the group trained with 1-2-8 and tested with 1-2-3. It was intended to repeat Zeiler's basic phenomenon and then to ask whether variations in training would affect the response in a way consistent with Stevenson and Bitterman's position. Our repetition of Zeiler's 1-2-8 : 1-2-3 condition with 28 4- and 5-year-old *S*s from day-care centers in Oakland, California, produced results shown in Row 2 of Table 1. Zeiler's results are shown in Row 1 for comparison. Zeiler's specifications for random spatial arrangements, placement of stimuli on the tray, correction procedures, and criterion (5 consecutive correct) were followed exactly. Despite the fact that Zeiler's procedure was duplicated as it is described and similar learning performance was observed, the authors were unable to obtain Zeiler's results in the test, and indeed the trend was in the opposite direction.

Zeiler (1963b) observes that the shift is most likely to occur if *S* attends to all three stimuli. Results in Row 3 are for 13 *S*s who were carefully instructed to look at all three stimuli before choosing. No appreciable change in behavior occurred. In a further effort to force *S* to attend to the difference between 2 and 8 in training, the sequence of stimulus arrangements was changed from one in which the various spatial arrangements occurred in random order to one in which 2 and 8 were always contiguous. Row 4 shows a possible slight effect from this change, but nothing resembling Zeiler's results. The authors have no satisfactory explanation of the inability to repeat Zeiler's results.

While the failure to repeat this critical result is unquestionably embarrassing to the AL interpretation, the failure to find an appreciable effect from attempts to ensure *S*'s learning of the 2-8 difference is also embarrassing for the position that transfer behavior is determined by the degree to which the training situation emphasizes absolute or relational cues. In an effort to show that what is learned during training determines the transfer behavior in this situation, the authors returned to a conventional, symmetrical middle-size problem in which training was on 1-2-3 and testing was on 2-3-4.

Gilinsky's (1955) work suggests that instructions can be used to affect the probability of relative and absolute responding, and Rock and Ebenholtz' (1959) study suggests that the relation between the background and the training stimuli is also a potent determiner. Consequently these two variables were selected for study.

After some trial and error, the authors realized that successful following of absolute instructions required that the background be particularly salient. To achieve this the background board on which the training and testing stimuli were displayed was modified by attaching three strips of white ¼-inch Masonite 6 inches apart. These strips, which ran from the bottom of the board (near *S*) to the top, were 2.37 inches wide—the same width as Stimulus 2.

On each training trial and on the test trial the three stimuli were placed on the 3 strips (random-placement orders). Stimulus 1 was smaller than the strip, Stimulus 3 was wider, but two sides of Stimulus 2 were always flush with the edges of the strip. During training, Stimulus 2 was reinforced. All *S*s were instructed to: "Find the right one—the one with the star under it." After the common initial instructions, *S*s were assigned to one of three different groups which received additional instructions. Group I (Relative) received additional instructions to: "Find the middle-sized one"; Group II (Absolute$_1$) was told: "See this one? The star will always be hidden under this one"; Group III(Absolute$_2$) was told: "Find the one that fits the board." These special instructions were not repeated on every trial during learning, and they were never repeated on the test trials.

TABLE 2

NUMBER OF SUBJECTS CHOOSING SMALL, INTERMEDIATE, AND LARGE TEST STIMULUS AFTER TRAINING WITH STIMULI 1, 2, AND 3

Instructional group	Test choice			
	N	S	I	L
I (Relative)	18	5	11	2
II (Absolute$_1$)	18	13	5	0
III (Absolute$_2$)	18	17	1	0

Note.—Test stimuli are 2, 3, and 4.

The results for these three groups shown in Table 2 demonstrate very clearly that *S*s' behavior in a transfer test after training on a middle-size problem can be controlled by selection of the proper combination of background and instructions ($\chi^2 = 18.23$, $df = 2$, $p < .01$; in computing χ^2 Columns 2 and 3 in Table 2 were combined). Relative instructions produce relative responding. Absolute instructions produce absolute responding. And absolute instructions that provide an explicit means of executing them (Group III) are almost completely effective.

SUMMARY AND CONCLUSIONS

Zeiler's AL account of earlier transfer results following intermediate size training is plausible, but because of its ad-hoc character it is no more conclusive than Stevenson and Bitterman's earlier explanation in terms of the relative importance of absolute and relational cues. Zeiler's own new evidence is at least as inconclusive: In six of seven transfer tests which permitted *S*s a choice between the stimulus predicted by AL theory and the one that was correct during training, *S*s chose the latter. In addition, the authors have been unable to obtain the change in test-stimulus preference following asymmetrical training that is predicted by AL theory. Our own data show that manipulation of both instructions and background have pronounced effects upon transfer behavior following intermediate size training. These effects, which can be attributed to the degree to which the training and testing situations emphasize the absolute or relational aspects of the stimuli, cannot be handled by current AL theories.

REFERENCES

BAKER, R. A., & LAWRENCE, D. H. The differential effects of simultaneous and successive stimuli presentation on transposition. *Journal of Comparative and Physiological Psychology,* 1951, **44,** 378–382.

GILINSKY, A. S. The effect of attitude upon the perception of size. *American Journal of Psychology,* 1955, **68,** 173–192.

JAMES, H. An application of Helson's theory of adaptation level to the problem of transposition. *Psychological Review,* 1953, **60,** 345–352.

RILEY, D. A., GOGGIN, J. P., & WRIGHT, D. C. Training level and cue separation as determiners of transposition and retention in rats. *Journal of Comparative and Physiological Psychology,* 1963, **56,** 1044–1049.

ROCK, I., & EBENHOLTZ, S. The relational determination of perceived size. *Psychological Review,* 1959, **66,** 387–401.

STEVENSON, H. W., & BITTERMAN, M. E. The distance effect in the transposition of intermediate size by children. *American Journal of Psychology,* 1955, **68,** 274–279.

ZEILER, M. D. New dimensions of the intermediate size problem: Neither absolute nor relational response. *Journal of Experimental Psychology,* 1963, **66,** 588–595. (a)

ZEILER, M. D. The ratio theory of intermediate size discrimination. *Psychological Review,* 1963, **70,** 516–533. (b)

(Received February 17, 1965)

Psychological Review
1966, Vol. 73, No. 3, 257–261

THE STIMULUS IN THE INTERMEDIATE SIZE PROBLEM

MICHAEL D. ZEILER[1]

Wellesley College

Riley, Sherman, and McKee maintained that a 2-process theory based on the learning of the absolute and middle size aspects of the stimulus was a better explanation of the intermediate size problem than the ratio theory. Various aspects of this contention were examined. The ratio theory appeared to be a more precise formulation than the 2-process alternative. In addition, no experiments with Ss below the level of adult humans that have used one training set have revealed that the middle size relationship can control behavior. Although most of the data were readily deduced by the ratio theory, there were results that required the assumption of precise absolute perception rather than the learning of ratios. Some alternative interpretations of the ratio theory were discussed, and a hypothesis based on 2 types of absolute learning was presented.

As an alternative to the ratio theory (Zeiler, 1963b), Riley, Sherman, and McKee (1966) offered a two-process hypothesis based on the learning of absolute and relational stimulus properties to explain the intermediate size problem (ISP). Coordinated with their explanation was the assertion that situational variables are important determinants of what is learned. The ratio theory, the two-process hypothesis, and the role of situational variables will each be discussed.

THE RATIO THEORY

The ratio theory states that what is learned in the ISP is the perceived size rather than the relational aspect (middle sizedness) of the positive stimulus. According to this theory, the perceived size of a stimulus is not defined solely by its absolute dimensions, but rather is described in terms of Helson's (1964) adaptation-level (AL) theory. This definition of the appearance of a stimulus as due to its relationship to the prevailing frame of reference or AL reveals the ambiguity in the way that the term "relational" has been used. Although the ratio theory asserts that relational principles are important because the perception of the size of a single stimulus depends on the other elements with which it appears, there is no reference to the relational concept of middle size. These two relational definitions of the stimulus have distinctly different meanings; the first deals with the specific stimulus as a percept, while the second pertains to abstract properties ("larger than," "smaller than," "middle sized") that transcend particular stimuli.

While Riley et al. (1966) asserted that the middle size relationship was learned in the ISP, they did not deny the importance of either the perceived size of the stimulus independent of its middle size quality or the role of frames of reference in size perception. What they did question was the concept used in the ratio theory that the frame of reference is changed by both the training and the test stimuli; they believe that such adaptation processes would occur only under very special circumstances. The idea of shifting levels is an integral part of AL theory that has been used by Helson (1964) to describe behavior in a wide variety of situations and was not a principle invented to explain the ISP. Whether or not such processes are the ordinary or the special case in the ISP or in any other area of psychology is a question to be settled by an experimental analysis rather than by argument. Pending such an analysis and in the absence of specific reasons for the rejection of this principle, it is difficult to understand why Riley et al. objected to this aspect of the ratio theory.

---

[1] I am grateful to Jerome Wodinsky for his critical reading and discussion of earlier drafts of this paper.

At any rate, the conception of perception as due to a shifting frame of reference does not exhaust the possibilities of what the ratio theory means. For example, suppose the subject (*S*) learns to choose a stimulus of a certain size in training and will continue to choose that same stimulus in the test. When transposition occurs, it is because *S* perceives no differences between the middle-sized stimuli of the two sets and is, therefore, responding to the "same" stimulus. The ratio theory shows how large a change is required in order that the middle size stimuli be just noticeably different from each other. That is, when the middle-sized member of the test set has the ratio most like that of the positive training stimulus, the two stimuli are perceived as identical; when another test-set member has the ratio closest to the positive training ratio, that element is perceived as being the same size as the positive training stimulus, and the middle-sized stimuli are not equivalent. The ratio theory, then, is primarily a way of talking about difference thresholds in the ISP rather than a statement about shifting perceptions.

Still another alternative is that the formulas used in the theory provide nothing more than a convenient and simple model to explain or to describe choice. The psychological meaning of this model is a separate issue.

In addition to accounting for the data of the ISP and predicting some new phenomena, the ratio theory has been specific enough so that its predictions could be tested and found wrong. So far, when it has been incorrect, explanations have been used that were compatible with the basic assumptions. In contrast, in the absence of statements that predict when either absolute or middle size learning will occur and will be demonstrated in the test, the two-process theory is ad hoc from the outset and has the dubious virtue of being impossible to prove wrong.

RELATIONAL AND ABSOLUTE LEARNING

Riley et al. asserted that the phenomenon reported by Zeiler (1963a) of the switch from a previously positive (Stimulus 2) to a previously negative (Stimulus 1) stimulus when the old positive stimulus was present in its original relationship (Train 1-2-8, Test 1-2-3) could be explained by the two-process hypothesis. They stated that *S* learned the nature and the magnitude of both the 1-2 and the 2-8 subrelationships in the training set and chose Stimulus 1 in the test because the 1-2 difference was held constant while the other pair was drastically altered (2-8 to 2-3). But why should *S* not choose either Stimulus 2 or 3 rather than Stimulus 1 if the 1-2 relationship played an important role in the learning of the training phase? Riley et al. maintained that such a subrelationship was learned in the training discrimination, but discounted its effect when discussing the choice in the test; in these terms, a switch to Stimulus 1 means that what was learned about the 1-2 relationship was either ignored or reversed on the test trial. Further research on the role of stimulus differences and subrelationships within a three-stimulus asymmetrical set is required before this explanation of the data can be entertained as an alternative to the ratio theory.

Setting aside considerations at the level of the relative explanatory and predictive merits of the two-process and the ratio theories, there is the question of a firm empirical basis for the two-process hypothesis. For the theory of Riley et al. to be plausible, there should be evidence that both relational and absolute learning occur in the ISP. The main problem in this respect involves the existence of relational learning in the sense that it is used in the two-process theory; that is, the idea that the abstract property of middle sizedness is learned.

Since transposition is deduced without reference to the learning of the middle size concept with the ratio theory, the mere occurrence of transposition does not prove that this relational principle is involved in the solution of the problem. What, then, *is* evidence that the middle size concept is important? A convincing demonstration would be a uniform gradient of perfect transposition with test sets of various degrees of difference from the training set. Such behavior, which would be best explained by reference to the middle size relationship, has never been obtained in the ISP. With the notable exception discussed below, transposition occurred only when the training and the test sets were very similar to each other. This is precisely what the ratio theory predicts.

Evidence for the learning of both the absolute and the middle size aspects of the positive training stimulus was provided by a study with adult humans (Zeiler, 1964) in which test choices were on either basis with any test set. On the first test trial, however, adults transposed when the training and the test set were very different from each other and responded to the previously reinforced absolute component when the two sets were similar. Here, in a study that revealed the simultaneous existence of both types of learning, the relational cue was used only when there was a large change between the sets. Since the adult behavior was opposite to that of children and infrahuman species who transposed with similar sets and chose the stimulus most similar in size to the positive training stimulus with more different combinations (see Zeiler, 1963b, for a review of the literature), there must be some differences in the underlying processes. Although the two-process hypothesis is not a compelling general explanation of the ISP since two processes are not needed to explain the test choices of children and infrahuman species, it has value when reserved to describe the behavior of adults.

SITUATIONAL VARIABLES

The author is in complete agreement with an emphasis on situational factors as determinants of what is learned in the ISP. But the following brief review shows that the effect of every situational variable that has been evaluated with a range of test sets was to change the amount of absolute learning that occurred. There is no evidence that any variable, with the possible exception of multiple training sets, displayed learning of the middle size relationship.

Recent findings revealed that behavior can be controlled completely by the absolute physical size of the positive training stimulus. Human adults given a successive discrimination (Zeiler, 1964) and pigeons given a simultaneous discrimination (Zeiler, 1965) always chose the test stimulus most similar in size to the positive training stimulus. Asymmetrical sets and sets composed of stimuli that were quite different from each other produced more absolute transfer than did those that were either more symmetrical or had smaller intraset differences (Zeiler, 1963b; Zeiler & Gardner, 1966).

While Riley et al. showed that instructions increased absolute learning (Group II), the choice behavior of the "relational instructions" group (Group I) was like that reported by Zeiler (1963b). Since in my experiment the same combination of sets was used but no special instructions were given, the relational instructions did not seem to alter the way in which the problem was learned without such directions.

Rudel (1957) found that a delay between the training and the testing phases resulted in transposition with a test set in which absolute choice occurred in an immediate test. The author has initiated research to determine whether the effects after a delay are due to middle size learning or occur for some other reason. If the middle size cue controls behavior several hours after learning, transposition should occur with test sets of various degrees of similarity to the training set. If, however, Rudel's (1957) data were due to the greater difficulty of differentiating the test from the training set after a delay, the specific test set with which response to the large or small stimulus occurs might be further removed from the training set than when testing is immediate. In the absence of such information, no conclusions about the role of delay are possible.

Riley et al. showed that a relevant background stimulus together with suitable instructions was another variable that produced almost perfect absolute learning. But the meaning of these results is ambiguous, as having *S*s match the positive stimulus to the background may have created a matching-to-sample situation which tapped different processes than does the ISP.

Certain combinations of two training sets did produce uniform transposition (Gonzalez & Ross, 1958; Zeiler & Paalberg, 1964). Since the use of two training sets in which the only common reinforced element is a certain relationship may extinguish the use of other cues, these experiments may have involved processes that do not play a part in the ordinary ISP. That learning of the middle size relationship can be demonstrated at all indicates that the potential for such learning exists, but it does not show that this potentiality is used in the discrimination of a single set.

WHAT IS LEARNED?

The absence of evidence for learning of the middle size relationship in any but the study with adults and the multiple training set experiments prevents the two-process theory from being a convincing explanation of the ISP. But, because of the growing catalog of situational variables that cause absolute learning, the ratio theory is also unconvincing.

The postulation of different kinds of absolute learning can account for all of the data. To the extent that contextual factors alter the perceived stimulus size, perception is not of the precise qualities of the physical stimulus. Such an imprecise perception of the stimulus size fits the ratio-theory model and explains transposition with similar training and test sets and absolute or random choice with less similar sets. Another type of absolute learning, resulting in absolute choices with all of the test sets, involves the exact perception of the individual stimulus size independent of and unaffected by changes in the context. These two types of perception, rather than being entirely different from each other, represent degrees of absolute perception; phylogenetic level and situational factors determine the point on the continuum of the precision of absolute learning that controls behavior in a given case.

The failure of Riley et al. to replicate Zeiler's (1963a) finding can be understood as a function of differences in absolute learning. My experiment indicated that the learning fit the ratio theory, while Riley et al. reported behavior that suggested a more precise perception of the size of the positive training stimulus. A few years ago, I reported difficulties in replicating this experiment and suggested that attentional factors might explain the discrepancies (Zeiler, 1963b). The isolation of the variables that operated in each of these studies to determine the type of learning that occurred would represent a substantial achievement.

REFERENCES

GONZALEZ, R. C., & ROSS, S. The basis of solution by preverbal children of the intermediate size problem. *American Journal of Psychology,* 1958, **71,** 742–746.

HELSON, H. *Adaptation-level theory.* New York: Harper & Row, 1964.

RILEY, D. A., SHERMAN, M., & MCKEE, J. P. A comment on intermediate size discrimination and adaptation-level theory. *Psychological Review,* 1966, **73,** 252–256.

RUDEL, R. G. Transposition of response by children trained in intermediate-size problems. *Journal of Comparative and Physiological Psychology,* 1957, **50,** 292–295.

ZEILER, M. D. New dimensions of the intermediate size problem: Neither absolute nor relational response. *Journal of Experimental Psychology,* 1963, **66,** 588–595. (a)

ZEILER, M. D. The ratio theory of intermediate size discrimination. *Psychological Review,* 1963, **70,** 516–533. (b)

ZEILER, M. D. Transposition in adults with simultaneous and successive stimulus presentation. *Journal of Experimental Psychology*, 1964, **68**, 103–107.

ZEILER, M. D. Solution of the intermediate size problem by pigeons. *Journal of the Experimental Analysis of Behavior*, 1965, **8**, 263–268.

ZEILER, M. D., & GARDNER, A. M. Intermediate size discrimination in seven- and eight-year-old children. *Journal of Experimental Psychology*, 1966, **71**, 203–207.

ZEILER, M. D., & PAALBERG, J. The effect of two training sets on transposition in adults. *Psychonomic Science*, 1964, **1**, 85–86.

(Received May 26, 1965)

Psychological Review
1966, Vol. 73, No. 3, 262–264

ON "ACTUARIAL METHODS AS APPROPRIATE STRATEGY FOR THE VALIDATION OF DIAGNOSTIC TESTS"

LEWIS R. LIEBERMAN[1]

Charles L. Mix Memorial Fund, Inc., Americus, Georgia

In a recent theoretical note, Sines (1964) has characterized old-strategy test validation as of the form: "If B (a particular type or class of behavior), then P (a particular test response or pattern)." He offers a new strategy of the form: "If P, then B." The present note shows that the old strategy (by using a control group) is really of the form: "P if, and only if, B." The new strategy is really of the form: "B if, and only if, P" and hence logically equivalent. The choice of strategy must be dictated by E's goals and hypotheses.

Sines (1964) has suggested a new "strategy" for the validation of psychological tests. He offers this new strategy as a substitute for and an improvement on the earlier methods. The old strategy, according to Sines, begins with some behavior (B) and attempts to find that test pattern (P) which could be inferred from that behavior. Sines symbolizes the logic of this approach as follows: "If B, then P." In the new strategy, we begin with P and attempt to find that B which can be inferred from *it* (in symbols: "If P, then B"). Sines is, of course, correct when he asserts that these two symbolic statements are not logically equivalent.

In favor of his new strategy, Sines makes two points. The first is that the new strategy is more like what clinicians really do (this accounts, he says, for why they are successful in real life, but not in validation studies). The second is that in fact a given B may manifest itself in several different Ps (the B to P mapping may be one to many), and this would make the search for the unique P futile.

The purpose of the present note is to set the analysis offered by Sines into a slightly larger context. The point to be made is that the new strategy is in fact logically equivalent to the old and that the choice of strategy should depend upon the purposes of the experimenter or upon his theoretical expectations.

LOGICAL CONSIDERATIONS

Let us begin with a closer analysis of the old strategy. The original derivation of the MMPI scales (e.g., Hathaway & McKinley, 1940) should serve as a good case study. The assumption behind this research was that there is a unique test pattern associated with a particular diagnostic category. This is clearly the old strategy ("If B, the behavior of the given diagnostic category, then P, the test pattern unique to the category"). Sines' fear that there may be several Ps for the category is not a problem here because if on a given statement some Hysterics, say, respond "true," and some respond "false," this item will not appear unique to B. The real worry here is that someone who is not Hysteric will turn up with a high *Hy* score.

The old strategists were aware of this danger, and to avoid it they used a control group. They were, after all, in search of items unique to the diagnostic category, and so they chose as items for the *Hy* scale those which significantly separated the Hysterics from their control subjects. In symbols, they tried to show: (a) "If B, then P"; and (b) "If non-B, then non-P." Given (a) and (b), then it is true that: "If P, then B"—the form of Sines' "new" strategy.

If a criticism can be leveled against the logic of these experiments, it is in the failure to have composed the control group of all possible non-B. If this had been done, then no item could have appeared on more than one scale. It is plain, however, that the answer to this criticism is not the strategy suggested by Sines, namely, checking to see whether Hysterics also show other patterns.

---

[1] The author is indebted to Richard M. Griffith, chief research psychologist, Veterans Administration Hospital, Lexington, Kentucky, for a critical reading of this manuscript.

Let us now examine closely the new strategy as exemplified by the model study reported by Sines. In this study, the records of patients exhibiting a certain MMPI profile (actually having profiles within an arbitrary distance of the 4′3 pattern) were checked for the behaviors reported in their case histories. If Sines were interested only in making an assertion of the form: "If P, then B," it would be sufficient merely to find a single behavior reported in all of the 4′3 patients' case folders. In actual practice, however, Sines employed a control group, and he looked for that behavioral item which those patients with a 4′3 pattern had *more than* those in the control group.

In symbols, Sines tries to show: (*a*) "If P, then B"; and (*b*) "If non-P, then non-B." Given *a* and *b*, then it follows that: "If B, then P"—the form Sines regards as old strategy.

The logical equivalence of the two strategies may also be demonstrated by noting that in Sines' study, the assignment of the factors P and B is completely arbitrary. The MMPI "test pattern" is as much behavior as is any other response or set of responses. Similarly, it does not strain the concept of test to regard case histories as tests. If we make this reassignment of letters, it is clear that this study is identical in form to Hathaway and McKinley's (1940) study.

The above analyses show that the old strategy by the use of a control group actually sought to show "P if, and only if, B." Similarly, through the use of a control group, the new strategy is of the form "B if, and only if, P." Logically, these two assertions are equivalent. The strategies differ only in where they begin: the old strategy beginning with some behavior for which a unique test pattern is sought, the new strategy beginning with a test pattern for which a unique behavior is sought.

EMPIRICAL CONSIDERATIONS

To illustrate the superiority of the new strategy, Sines conducted a second study which presumably shows how an old strategist would have conducted his first study. The results were negative: Only 1 person of 31 who were judged assaultive did have the 4′3 MMPI pattern. This finding is not proof that the two strategies yield different results. Since we have shown that they are logically equivalent, this second study is simply a contradictory finding.

There are several aspects of Sines' procedures which might account for contradictory results. First, Study I was never cross-validated. Second, it should be noted that Study I did not show "If non-P, then non-B" exactly (since 40% of the non-P showed B); therefore, the first or the second study may be a chance finding. Third, there were several other Bs (see Sines, 1964, p. 521, Table 1) which should have been investigated also. Fourth, there is some question concerning the test for B in the two studies. In Study II, persons were selected who had either "threatened to kill or had seriously assaulted someone." In Study I, however, the criterion was "history of hostile or aggressive behavior." It should be noted that none of the 4′3 patients who met that criterion was "felt to be dangerous" or "had a violent temper"; whereas six of the non-4′3 patients (control group) were felt to be dangerous, and three were judged to have a violent temper. This latter point raises the general question of the adequacy of a social history for this kind of research.

In further support of the superiority of the new over the old strategy, Sines presents a hypothetical example. In this example a particular B, assaultive behavior, manifests itself in several different test patterns. We are given the following invariances: Of 100 persons showing B, 40 have test pattern P_1, and 15 each have test patterns P_2, P_3, P_4, and P_5. We have somehow isolated only test pattern P_2 and are interested in finding the relation of this pattern to B. Sines contends that if such a relation existed, the old strategy would fail to reveal it. This is not correct, however.

Using the old technique, we would collect cases of B and examine the tests of these individuals for the presence of P_2. We would also randomly select a control group of non-B cases from the same population. If we collected 100 cases of each, they would fall into four discrete groups: B, P_2; B, non-P_2; non-B, P_2; non-B, non-P_2. According to the conditions of the problem, the numbers in each of these classifications would be, respectively, 15, 85, 0, 100. A chi-square on these data woud be equal to 16.2, significant beyond the .001 level, revealing a definite relation between B and P_2.

It must be admitted that the new strategy which would have had us collect cases of P_2 and non-P_2 and examine these for B would have yielded a more efficient result since the number of cases required would have been smaller. However, rarely does one wish merely to relate one test pattern to one behavior. More frequently the attempt is made to relate one (or several) behavior(s) to several test patterns. In the model study reported by Sines, several behaviors were to be related to the one test pattern.

CHOICE OF STRATEGY

Since the two strategies are logically equivalent and should produce the same empirical results, the choice of strategy must depend on the needs of the experimenter. The test maker in his thinking must begin with B. He wants by definition to create a situation (the test) in which B will manifest itself either directly or indirectly. P is the indicator, the sign of B. His rationale explains why this relation exists. A "P, B" relation which just exists leaves a theoretical gap in his work.

For all the examples given by Sines, the MMPI, Rorschach, Thematic Apperception Test, and Draw-A-Person Test, rationales exist. Sines' own MMPI study begins with a pattern which was evolved from old-strategy thinking. Given just the responses to the 566-item MMPI, it is difficult to imagine how one would begin to isolate patterns which would be relevant to socially significant behavior. Were this even possible, the question of "why" would nag and demand attention.

Once a test has been devised (and perhaps only then), it may be possible for a user to develop a hypothesis for which the new strategy would be appropriate.

Furthermore, it is unlikely that the test users, the experts mentioned by Sines who can infer a personality structure from a Rorschach or MMPI profile, do so by sorting mentally through all the profiles they have seen. More likely, their thinking represents a combining of the various behavior tendencies which they see indicated by the elements of the profile.

Sines' analysis should serve to alert test validators to the possibilities of a given behavior pattern manifesting itself in several test patterns. The other side of the coin should not be overlooked, to wit, that a single test pattern may indicate several kinds of behaviors. The choice of strategy must be dictated by the experimenter's goals and hypotheses.

REFERENCES

HATHAWAY, S. R., & MCKINLEY, J. C. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology,* 1940, 10, 249–254.

SINES, J. O. Actuarial methods as appropriate strategy for the validation of diagnostic tests. *Psychological Review,* 1964, 71, 517–523.

(Received March 1, 1965)

VOL. 73, NO. 4 JULY 1966

PSYCHOLOGICAL REVIEW

LAWFUL DISORGANIZATION:

THE PROCESS UNDERLYING A SCHIZOPHRENIC SYNDROME[1]

WILLIAM E. BROEN, JR.
University of California, Los Angeles

AND

LOWELL H. STORMS
Neuropsychiatric Institute, UCLA Center for the Health Sciences

Schizophrenics tend to exhibit deviant behavior in a number of different areas. This paper attempts to account for schizophrenic association and conceptualization disturbances, delusions, hallucinations, and regression in terms of a common process involving partial randomization of response hierarchies. Because of a common underlying process, these different areas of schizophrenic behavior are expected to have similarities in terms of how they differ from normal behavior, and conditions under which behavioral deficit will be observed. A formal account which involves the facilitating effects of increased arousal on competing responses, when appropriate responses are at response-strength ceiling, is used to discuss similarities which are expected across content areas. The implications from the account appear to integrate research findings and clinical observations concerning the different kinds of schizophrenic abnormalities discussed.

This paper is concerned with the question, "Why do different behaviors which are thought of as 'schizophrenic' go together?" Our purpose is to present a possible explanation for the clustering of some schizophrenic behaviors, basing the explanation on a process which is seen as underlying different areas of schizophrenic behavior. The discussion is not intended as a definitive account of all behaviors which occur in persons diagnosed as schizophrenic, partly because much of what will be said requires additional support, and partly because our goal is to emphasize only certain related aspects of the behaviors characteristic of this heterogeneous nosologic category, schizophrenia. We will emphasize behavior changes which are a part of a process which would be expected as a direct result of increased arousal which facilitates the various response tendencies evoked in a situation. These behaviors include the type of disturbances in thought association and conceptualization which have been considered basic or of major importance in schizophrenia (Bleuler, 1950; Cameron, 1939, 1944; Chapman, 1956, 1958, 1961; Meehl, 1962; Payne, 1961). We will also discuss regression, delusions, and hallucinations as related behaviors. We will not em-

---

[1] Part of this paper was written while the senior author was a special consultant on a research project supported by Grant No. R-61-2-20 from the California Department of Mental Hygiene.

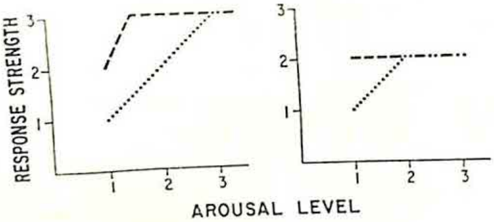

FIG. 1. Strengths of dominant (--) and competing (....) responses as a function of arousal level and level of response-strength ceiling. Level of response-strength ceiling is 3 in the left figure, 2 in the right.)

phasize defensive behaviors which serve an anxiety-reducing function. There is little doubt that defensive behaviors are used by schizophrenics. However, defensive behaviors will be seen as probable, but not necessary, correlates of the process which is the focus of this paper. This focus follows two complementary lines of thought: (*a*) that behavioral deficit in schizophrenia is the result of arousal or anxiety (see literature reviewed in Distler, May, & Tuma, 1964; Lynn, 1963; Mednick, 1958), and (*b*) that behavioral deficit may be related to arousal apart from possible defensive or avoidance functions of the behavior (Fish, 1961; Freeman, 1960; Malmo, 1959).

The basic theoretical account of the arousal-related process to be used here has been presented previously (Broen & Storms, 1961). In this account, a stimulus situation evokes habitual tendencies to respond which vary in strength depending on amount of previous training. Habit strength interacts multiplicatively with arousal (drive) level to yield response strength. When more than one tendency to respond is evoked in a situation, the probability of the dominant response is an increasing function of the differences in response strengths between dominant and competing responses. A response-strength ceiling which is lower than the product of maximum arousal level and maximum habit strength is also postulated. In this account, if arousal level and habit strength of the dominant response are sufficient to raise the strength of the dominant response to its ceiling, further increases in arousal can only facilitate competing responses. The result is a decrease in the relative dominance of the dominant response, with increased probability of competing responses.

The basic relationships in the account are illustrated in Figure 1. Note the figure on the left first. This figure is intended to illustrate alternate responses which are evoked in different strengths (e.g., the dominant response "light" and competing response "night" evoked by the stimulus word DARK in a word-association task). When arousal level is increased beyond the level where the strength of the dominant response reaches ceiling, the hierarchical order of dominant and competing responses tends to break down. With increasing arousal, the tendency is for alternate responses to become more equal in strength and probability of occurrence (Arousal Level 3 in Figure 1).

Normal, appropriate behavior is characterized by dominance of appropriate responses. Thus in an appropriately responding person, an increase in arousal which decreases the frequency of dominant responses usually reduces appropriateness of behavior. This decreased appropriateness will be of a particular type. With decreased hierarchial dominance of appropriate over competing responses, behavior will be characterized by increased fluctuation between appropriate and competing responses. This type of disorganization of normal response hierarchies seems to be common to each of the content areas of schizophrenic behavior we will discuss.

As is illustrated in Figure 1, note that the extent of hierarchical disorganization is a function of three variables: (*a*) the strength of dominant and competing responses, (*b*) the level of arousal, and (*c*) the level of response-strength ceiling. If response strengths of dominant and competing responses are low, an increase in arousal may increase the dominance of the appropriate response (as in the increase in arousal from 1 to 1.5 in the left figure). Also, if competing responses are not present, there is no decrement in the relative strength of the appropriate response regardless of arousal level. The effect of individual differences in ceiling level can be seen by comparing the left and right figures. A low ceiling leads to disorganization of a response hierarchy under lower levels of arousal. It is important to note that because of differences in level of ceiling we expect the same type of disorganized behavior to occur at different arousal levels in different persons. While we expect similar changes in behavior to occur across individuals as a function of intraindividual variations in arousal, a specific level of arousal will not be assumed to be related to a specific amount of disorganization of response hierarchies. Not all schizophrenics exhibiting the type of behavior to be discussed here will be expected to be characterized by high levels of arousal.

Some additional comments will aid in translating the account into experimental predictions. The term arousal is used to indicate a dimension of relatively diffuse cortical and physiological activation which may be measured by changes in peripheral physiological indicants such as increases in heart rate, blood pressure, and respiration rate. Extended discussion concerning the concept of arousal or activation as we use the terms can be found in Carlton (1963), Gellhorn (1964), Malmo (1959), Pinneo (1961), and Schnore (1959). The studies we will cite as having varied arousal experimentally use independent variables which are similar to variables which have been shown to lead to the general physiological changes indicative of arousal. A particularly good example of such a variable is induced muscular tension (Pinneo, 1961). Variation in anxiety or psychological stress is also seen as related to arousal variation (Schnore, 1959; Spence, 1958b), although caution is needed in the interpretation of studies using specific methods of varying anxiety which have not been related to physiological indicants of arousal.

We have used three measures of habit strength in making predictions for our own research and in interpreting other research. In experiments in which new S–R connections are trained, the habit strength of a response is assumed to be an increasing function of the number of occurrences of that response during training. Habit strength is also assumed to vary inversely with the amount a stimulus differs from a training stimulus when the subject is familiar with the stimulus alternatives in the experiment. However, these definitions are of little use in many of the studies we will cite because these studies deal with responses with unknown learning histories. In studies using stimuli and responses which are commonly used, for example, associations to familiar words, we will assume that the habit strength of the dominant response is at its maximum, and the average habit strength of a competing response for groups of schizophrenics is a function of the probability of that response as observed in normals.

Response-strength ceiling effects in experiments are predicted on the basis of two assumptions. First, schizophrenics are assumed to have lower average response-strength ceilings than nonschizophrenics. As a result, the response hierarchies of schizophrenics should become disorganized at lower levels of arousal. This assumption is consistent with data suggesting that normals may operate efficiently within a broader range of arousal than schizophrenics (Lynn, 1963; Venables, 1960). Second, we will assume that, in a group of persons who exhibit schizophrenic behavior in hospital environments, the stress of standard experimental situations is sufficient to lead to ceiling effects when familiar associations are tested. With this level of arousal and lower ceilings, schizophrenics should exhibit more of the type of disorganization expected from response-strength ceiling effects than nonschizophrenics.

The implications of the present account overlap in some respects hypotheses proposed by Chapman (1958), Payne (1961), Meehl (1962), especially in his second neurological model, Weckowicz and Blewett (1959), and Stilson and Kopell (1964) in their neurophysiological thinking. The account also has a degree of similarity to Russian thinking concerning a reduction in internal inhibition in some schizophrenics which would selectively facilitate nonreinforced responses and responses to weak stimuli (Lynn, 1963). There is also a great deal of similarity to the neurophysiological theory of schizophrenia proposed by Fish (1961).

Because of common background in extensions or modifications of concepts from Hullian learning theory, this account is also similar in some respects to an account of inappropriate behavior in terms of competing responses raised above threshold by increased drive (Spence, 1956, 1958a; Taylor & Spence, 1952). This threshold theory was used by Mednick (1958) in a very stimulating account of some aspects of schizophrenia. However, the accounts differ on the basic point of how increased arousal leads to inappropriate behavior and thus also differ in predictions regarding schizophrenic behavior (for discussions of the accounts see Broen & Storms, 1961; Broen, Storms, & Schenck, 1961; DeMille, 1959; Hill, 1957).

It may be useful to note a basic difference between the ceiling account and the threshold account. When response-strength threshold factors, rather than ceiling effects, are seen as leading to decrement in a dominant response, increased arousal is expected to lead to decrement only when response strength is low. In the usual situation where a response which is appropriate has higher habit strength than a competing response, threshold effects would lead to the following:

> . . . at very low levels of drive and training a relatively weak-drive group would be expected to perform at a higher level than a somewhat stronger-drive group. On the other hand at higher levels of training and drive the reverse picture would hold; that is, the performance of the higher-drive group should be above that of the relatively weaker-drive group [Spence, 1958a, p. 92].

In contrast, response-strength ceiling effects would lead to dominant-response decrement primarily under higher levels of drive and higher levels of strength of dominant and competing responses. The following discussion of some aspects of schizophrenic behavior is based on consideration of response-strength ceiling effects.

Association Disturbance

Both clinical and research literature suggest that schizophrenics have different associations in their thought processes or, on a more overt level, different verbal associations from normals or other psychiatric patients (Deering, 1963; Sommer, Dewar, & Osmond, 1960; Storms & Broen, 1964). Looking at the particular nature of this difference, it seems that the schizophrenic's associations are not disconnected, highly irrelevant thoughts. Instead, schizophrenic associations are characterized as "unprecise approximations" which "cannot be dismissed as either 'irrelevant' or 'incoherent' for they are certainly about the general subject and loosely related to one another [Cameron, 1944, pp. 53–54]." Bleuler (1950) described the deviant but related associations another way, characterizing some schizophrenic thinking as "ideas of a certain category . . . thrown into one pot, mixed, and subsequently picked out at random [p. 16]." As a more specific example of schizophrenic "looseness of association," Arieti (1955, p. 4) considers the schizophrenic disorder to be implicit in the answer, "White House," to the question, "Who was the first president of the United States?" Burstein's (1961) study comparing schizophrenics and normals on selecting synonyms from a list of words also suggests that the primary verbal difference between schizophrenics and normals is the schizophrenics' overuse of associations which are inappropriate but related to the stimulus, such as antonyms and homonyms.

This particular form of association disturbance seems to be consistent with the kind of behavior expected under high arousal within the account used in this paper. As has been noted, an increase in arousal beyond the level necessary to raise a dominant response to maximum strength will tend to equalize the response strengths of dominant and competing responses. Thus, the expected form of disturbed behavior would be characterized by increased randomization among the responses in response hierarchies (Bleuler's, 1950, "ideas in a certain category") with a consequent increase in the probability of those behaviors which are at relatively low strength in normal response hierarchies. The increase in within-hierarchy, associated-but-deviant responses seems much like the descriptions of the characteristics of associative disturbance in schizophrenia mentioned above.

If schizophrenic associations can be primarily understood as associations under high arousal, the formal account has other implications in addition to the expectation that deviant responses will be loosely relevant. The partial equalization of response strengths of alternate responses under increased arousal also implies an instability of associations. In addition, in schizophrenics, instability of associations should increase with increased arousal level. Greater instability of associative responses in schizophrenics than in normals was found on repeated association tasks by Sommer et al. (1960) and Storms and Broen (1964). This associative instability in schizophrenics does seem to increase under stress conditions, such as time stress and in response to words rated as anxiety producing, and the increase in instability under stress tends to be greater in schizophrenics than in neurotics and normals (Storms & Broen, 1964). The latter result was interpreted as being due to schizophrenics having greater arousal reactivity to stress conditions, or having lower response-strength ceilings, and therefore being more likely to show the response-strength ceiling effects discussed above.

As was noted previously, the account used here also implies that schizophrenic association disturbance should be greatest on tasks where competing responses are present and have relatively high habit strength. This implication is consistent with data indicating that, relative to normals, schizophrenics exhibit greater deficit on tasks requiring the use of words with multiple meanings than on tasks using single-meaning words (Faibish, 1961). In another relevant study, Blaufarb (1962) compared schizophrenics and normals on the meanings associated with proverbs, with the ambiguity of the proverbs varied by two methods of presentation (i.e., presenting proverbs singly, and presenting three proverbs having the same meaning in a group, with the group method assumed to provide a more precise stimulus which would reduce the number of inadequate associations). The performance of schizophrenics was significantly poorer than that of the normals only under the more ambiguous method of presenting proverbs, again suggesting that schizophrenics' associations are most deviant on tasks where competing associations are present.

Paired-associate learning tasks have also been used to investigate the hypothesis that schizophrenics are especially deficient on tasks where competing responses are present in strength. The results are inconsistent but apparently tend to support the hypothesis (Kausler, Lair, & Matsumoto, 1964; Spence & Lair, 1965, support the hypothesis, while Spence & Lair, 1964, report negative results). However, it is difficult to interpret the results of these studies in terms of the relative decrement in dominant response expected from the present account. This is because during training the correct response is not dominant at first and then becomes dominant, but the data are overall measures of performance which do not take variation in dominance into account.

In summary, there is some support for the hypothesis that schizophrenic associations may be influenced by the type of disorganization of response hierarchies expected from ceiling effects: (*a*) Schizophrenic deviant associations tend to be loosely associated with the stimulus; (*b*) schizophrenic response hierarchies tend to be unstable, with this instability increasing under stress; and (*c*) schizophrenic association deficits tend to occur primarily in situations which elicit multiple associations.

CONCEPT DISTURBANCE

Certain changes in conceptual responses are also expected as a result of increased arousal. By "conceptual responses" we mean responses which are appropriate only to specific classes of objects (e.g., using the label "birds" only to designate birds). The same task and arousal conditions which were relevant to the occurrence of "looseness of association" are relevant to concept disturbance. When appropriate conceptual responses are dominant and competing responses are also evoked at high strength, increased arousal may increase the frequency of the competing conceptual responses. As an example, consider a task which has been used in research on concept disturbance (Chapman, 1961). A subject is given cards on which the names of different objects are printed, and he is instructed to select the cards which refer to birds. Some of the cards designate flying insects. When the subject looks at a card designating a flying insect, the response which is appropriate to bird cards (selection) will be evoked in some strength due to mediated generalization. This generalized response tendency will compete with appropriate conceptual responding. Increased arousal which is sufficient to lead to response-strength ceiling effects would increase the frequency of competing responses. Some insect cards would be selected as bird cards. This is an example of an overly broad or overinclusive use of a conceptual response which may result from increased arousal.

Increased arousal may also result in a too narrow or overexclusive use of a conceptual response. Some objects which should be included in a class of objects may be similar in some respects to objects which do not belong in a class. Such objects would evoke competing tendencies not to include them in their appropriate category. With sufficient increase in arousal, these competing tendencies toward inappropriate exclusion would be exacerbated. Note that overexclusive and overinclusive responding are not mutually incompatible. If competing response tendencies to include certain inappropriate objects in a class and to exclude certain appropriate objects are both present, increased arousal may increase both types of conceptual errors.

Deviant breadth of concepts is a commonly noted aspect of schizophrenic behavior. Consistent with the above expectations, schizophrenic concept performance has been characterized both as overexclusive ("concrete") and overinclusive (see literature reviewed by Chapman, 1961; Payne, 1961). Usually studies have found schizophrenic concepts to be either overexclusive or overinclusive, with more recent studies emphasizing overinclusion. A few studies have found schizophrenic conceptual performance to be both overexclusive and overinclusive (Chapman, 1961; Chapman & Taylor, 1957; Zaslow, 1950). In some contrast to the rationale of the present discussion it has been suggested that type of conceptual error in schizophrenics is a function of type of subjects, with organic pathology in "process" schizophrenics leading to overexclusive concepts (e.g., Tutko & Spence, 1962). Subject and task variables may both be important in determining type of conceptual disturbance. Cortical damage in some schizophrenics may lead to overexclusion, while an arousal process in others leads to overexclusion or overinclusion, depending on task conditions.

TABLE 1

MEAN ERRORS ON CONCEPT-SORTING TASKS FOR SCHIZOPHRENICS AND NORMALS

Task	Overinclusion errors		Overexclusion errors	
	Normal	Schizophrenic	Normal	Schizophrenic
1	.12	1.61	1.52	2.61
2	.33	2.08	2.94	8.06
3	.42	6.85	.94	2.40
4	2.12	10.57	.61	1.39

Note.—Adapted from Chapman (1961).

Within the present account, increased arousal will increase a particular kind of error only in a task which evokes a tendency toward that error in some strength. To the extent that differences between schizophrenics and normals are due to the arousal process, this means that schizophrenics should make the same types of conceptual errors as normals, except that schizophrenics should make more of these errors. Also because exacerbation of errors under arousal depends in part on the strength of error tendency evoked by a task, the frequency of errors across tasks should tend to be correlated for schizophrenics and normals. Evidence on the above points is provided in the study we have used to provide a research example (Chapman, 1961). Chapman used four kinds of conceptual card-sorting tasks and recorded both overinclusive and overexclusive errors. Relevant data from Chapman's study are presented in Table 1.

Note several things in Table 1. First, both overinclusion and overexclusion errors are made by schizophrenics and normals, with schizophrenics making more of both types of errors. Also, in three of the four tasks the dominant error direction (overinclusion versus overexclusion) is the same for both schizophrenics and normals. In addition, note that for overinclusion and overexclusion errors the error frequencies for schizophrenics and normals have 1.0 rank-order correlation across tasks. It may also be of interest to note that in a different content area, syllogistic reasoning, Gottesman and Chapman (1960) found that hierarchies of reasoning errors within tasks were similar for schizophrenics and normals, the difference consisting of a partial randomization of response hierarchies in the schizophrenics. As expected from the paradigm used here, the frequency of a particular type of conceptual error does seem to be a function of the strength of that error tendency in normal response hierarchies. The sugestion that schizophrenic error patterns seem to be exacerbations of normal error patterns (with supporting evidence) has also been made in other papers by Chapman (1958) and by Fey (1951).

We have emphasized the relationship of schizophrenic conceptual errors to the strength of error tendencies in normal response hierarchies. In the paradigm being used here, the strength of an appropriate dominant response also plays a role in determining whether or not competing inappropriate responses will increase under increased arousal. As was seen in the discussion of Figure 1, response-strength ceiling effects leading to an increase in competing responses are most likely when dominant response strength is high. Concept research with schizophrenics has usually not emphasized strength of appropriate conceptual responses as an independent variable. The stimuli have tended to be familiar, supposedly eliciting appropriate responses of high strength. Stimulus-generalization studies may provide relevant evidence concerning variation in strength of dominant responses as related to errors. We have some hesitation about citing stimulus-generalization studies in the present context because concepts are usually based on mediated generalization instead of primary stimulus generalization. However, as suggestive evidence, note that in stimulus-generalization tasks, schizophrenics as compared to normals do tend to exhibit reduced frequency of a dominant response primarily where the response is strong—at the training stimulus (Broen & Storms, 1964; Dunn, 1954; Garmezy, 1952; Knopf & Fager, 1959). Recent evidence suggests that the same phenomena may occur in the area of mediated generalization (Chapman & Chapman, 1965; Peastrel, 1964). In these studies, schizophrenic deficit in responses which depended on mediated similarity occurred primarily to stimuli where mediated similarity was greatest.

Consideration of the importance of task variables in determining type of concept disturbance in schizophrenia suggests that the recent emphasis on schizophrenics' overinclusive concepts in the research literature may be a function of task conditions. We hypothesize that overinclusive and underinclusive errors would occur with equal frequency in acute schizophrenics if task conditions were equated. Equating task conditions would involve equating appropriate response strengths and competing response strengths for stimuli where inclusive and exclusive responses are appropriate, and having equal numbers of both classes of stimuli in the study (equal numbers of objects which would be correctly included and excluded).

An additional point is that if schizophrenic conceptual errors are responses which are low in normal response hierarchies, schizophrenic conceptual errors should be associated responses rather than irrelevant responses. Referring back to the research example, this means that if there is overinclusion in selecting objects which are birds, the specific overinclusions would tend to be flying insects rather than irrelevant objects such as articles of clothing. The research evidence indicates that schizophrenic conceptual errors do tend to be associated errors rather than irrelevant errors (Chapman, 1958, 1961; Chapman & Taylor, 1957).

The arousal paradigm also implies that schizophrenics should show more conceptual deficit under increased arousal (except when dominant appropriate responses have low strength—see Figure 1—or when an aversive stimulus is used specifically to provide new information or as an incentive—see Cavanaugh, 1958). The most relevant data we know on this point are again from stimulus-generalization studies. Broen, Storms, and Goldberg (1963) tested the generalization of a motor response under different amounts of induced muscular tension (arousal). Increased arousal led to a decrease in the response at its training stimulus (overexclusion) and a facilitation of the response to generalization stimuli (overinclusion). In another relevant study, Broen and Storms (1964) trained schizophrenics to press a lever in one direction in response to one category of pictures (tree with branch angles in a specified range) and to press the lever in the other direction in response to another category of pictures (different branch angles). Lengthy training and probabilistic reinforcement were used to insure that both appropriate and competing responses had high habit strengths. An increase in induced muscular tension reduced the number of appropriate responses and increased the number of competing responses. In accordance with an hypothesis of higher response-strength ceilings in normals, hospitalized nonpsychiatric subjects showed no significant decrement in appropriate responses or increase in competing responses. These two studies are most directly relevant to the effect of arousal on conceptual responses when concepts are defined in terms of a range of stimulus similarity. In these studies the evidence is consistent with the implication from the account that under certain boundary conditions increased arousal tends to increase competing conceptual responses at the expense of appropriate conceptual responses. We expect that increased arousal has similar effects on responses to other types of concepts, such as conceptual groupings determined by functions of objects, but research which would confirm this generalization is certainly needed.

DELUSIONS

We expect that transitory and relatively unstructured delusions will accompany the types of association and concept disturbances we have been discussing. When a person's categorizations are underinclusive and overinclusive, the data which he sees as relevant to a problem will be deviant. In drawing conclusions some relevant data will not be used, and loosely associated data, which are normally excluded as unimportant, will be seen as quite relevant. Delusional conclusions will result. These delusions will be transitory and unstructured because the underlying process involves fluctuations between appropriate and competing associations. It is important to note that increased arousal should lead to a syndrome which includes only transitory and unstructured delusions. Stable delusions may develop later if a particular delusion reduces anxiety. Such stability of thought implies that the current level of arousal is not sufficient to lead to response-strength ceiling effects.

Cameron's (1951) conclusions concerning delusions are quite similar to the implications of the present account. He distinguished two stages in the formation of a delusion—an initial diffuse and unstructured phase and a later restrictive phase. The unstructured delusions are related to anxiety through a causal chain like that in the present account, namely the effect of anxiety in lowering the "relevance-to-irrelevance" ratio of the information on which conclusions are based. The development of structure and stability would occur later if the delusional thinking comes to serve a defensive function.

Relevant research information is sparse but supportive enough to suggest that further investigation based on the implications of the present account may be useful. As expected from the postulated relationship between concept disturbance and delusions, recently admitted schizophrenics who have delusions are significantly more overinclusive in their thinking than either nondeluded schizophrenics or nonschizophrenic patients (Payne, Caird, & Laverty, 1964). This relationship seems to be a function of chronicity and may be reversed in chronic deluded schizophrenics if delusions are the single prominent symptom as would be expected when a delusion serves a defensive function (Silverman, 1964).

HALLUCINATIONS

An arousal-induced disorganization of response hierarchies may also underlie the occurrence of some types of hallucinations. When internal or external cues evoke competing perceptual responses, increased arousal may lead to a relative decrease in the dominance of appropriate perceptual responses, thereby increasing inappropriate perceptions.

The hallucinations which would result from this breakdown in normal response hierarchies would tend to be disorganized and fragmented mixtures of appropriate and inappropriate perceptions. These hallucinations should also increase with increased arousal. In addition, because the hallucinatory perceptions are normally evoked as low-strength competing responses, the misperceptions should have some association to the stimulus situation in which they occur. We will briefly discuss each of these implications in turn.

The expectation that some schizophrenic hallucinations should be disorganized mixtures of appropriate and inappropriate perceptions is different from the rather organized and meaningful reports of hallucinations which are usually presented in abnormal psychology texts. However, there is a fair amount of clinical evidence suggesting that schizophrenic hallucinations are often, even usually, such fragmented mixtures. For example, Kraepelin (1919) and Bleuler (1950) both report rather disorganized and fragmented content in schizophrenic hallucinations. Bleuler emphasizes the varied and abrupt nature of hallucinations and observes that sensory deceptions may occur along with a correct observation of the same event. Discussing hallucinated voices Bleuler (1950) states that: "As a rule . . . the patients hear short sentences or single words which in themselves need not always make sense. It is the patient who usually imputes some meaning to the words or sentences [p. 99]." This meaningful organization of misperceptions may be a secondary process which is learned because it is anxiety reducing. As Kraepelin (1919) notes: "They [hallucinations] are almost never wanting in the acute and subacute forms of the disease . . . at the beginning these are usually simple noises . . . and then develops . . . the hearing of voices [p. 7]."

The hypothesized relationship between increased arousal and hallucinations seems at first to be contrary to the discussions which have emphasized the causal role of low arousal, primarily being concerned with dreams during sleep (see Evarts, 1962). However, Harris (1959) has emphasized the point that this low-arousal process is different from the process leading to hallucinations in schizophrenia, a suggestion which is supported by physiological differences between dreaming and schizophrenia (Rechtschaffen, Schulsinger, & Mednick, 1964). Other writers point out that high levels of arousal or anxiety may also be associated with hallucinations (Fish, 1961; West, 1962; Zuckerman, Albright, Marks, & Miller, 1962). A high-arousal explanation of schizophrenic hallucinations also seems consistent with differences in frequency of different types of hallucinations. For example, hallucinations seem to have more stressful than nonstressful content (e.g., accusing voices). Stressful thoughts would be expected to lead to higher levels of arousal than nonstressful content, increasing the likelihood of lack of discrimination of the source of the stimulation. We might also consider why unusual perceptions in schizophrenia are so frequently concerned with body sensations (note the MMPI *Sc* scale content) and are more frequently auditory than visual (Malitz, Wilkens, & Esecover, 1962). The explanation may lie in different arousal-producing properties in different sensory modalities. If a particular kind of stimulation leads to more arousal, these stimuli will be more usually misperceived. In line with the dominance of body and auditory hallucinations in schizophrenia, Bernhaut, Gellhorn, and Rasmussen (1953, p. 33) state that ". . . sensory stimuli listed in decreasing order of ability to produce an arousal reaction are: (i) pain (most effective); (ii) proprioception; (iii) auditory stimulation; (iv) optic stimulation (least effective)."

The suggestion that hallucinatory perceptions should be perceptions which are associated with the situation in which they occur also seems consistent with clinical observations and the results of a few relevant experiments. The association is most obvious when the misperception is simply an inaccurate localization of stimulation. For example, patients' own thoughts appear to be spoken out loud, and actual external sounds may be perceived as coming from within (Bleuler, 1950). The view that hallucinations are associated responses also receives some experimental support from studies indicating that auditory hallucinations are more likely in the presence of audible sound (Harris, 1959; Riss, 1959), and from Lindsley's (1963) finding that "in general, the more similar the stimulus was to a human voice, the higher the probability that a given patient would 'talk to it' [p. 295]."

REGRESSION

In normal adults, those responses which are dominant in a situation tend to be appropriate to that situation. Competing responses may also be evoked at some strengths. These competing responses may be of at least two types: (*a*) responses which had been learned earlier but are no longer appropriate, and (*b*) responses which are generalized from similar situations. Increased arousal which is sufficient to equalize dominant and competing responses at ceiling strengths will therefore increase the proportion of responses which had been learned earlier in the situation (or related situations). In the sense that the probability of old responses is increased, regression is expected as one aspect of the syndrome of behaviors expected under high levels of arousal. However, the expected behavior is not simply a return to childlike patterns of behaving. The form of regression would be a relatively unstructured mixture of appropriate, generalized, and regressive responses.

There is some evidence that schizophrenic regression is not a simple return to childlike behavior patterns (see literature reviewed in Chapman, Burstein, Day, & Verdone, 1961), and that the lack of subordination of responses which are normally more hierarchically ordered is the form of regression which occurs in schizophrenia (Cameron, 1939). Cameron's conclusions from an object-sorting task he used to study regression provide a good description of the type of regression which occurs in schizophrenia. He observed that the schizophrenics were ". . . unable to organize or subordinate the events occurring simultaneously in their organism—perceptual, memorial, imaginal [p. 269]." While such behavior may have some childlike qualities, it would seem to be more aptly described as a decrement in the normal hierarchical ordering of alternate response tendencies.

Much of the research suggesting that regression can be induced by increased arousal has been done with infrahuman subjects (see Kleemeier, 1942). However, there is considerable clinical literature and some experimental evidence suggesting that increased stress or arousal is related to regression in humans (e.g., Barker, Dembo, & Lewin, 1941; Levin, 1965; Noyes & Kolb, 1958; Rodnick, 1942; Sadler, 1953).

Levin's (1965) study provides an especially rigorous test of a relationship between increased arousal and regression implied by the paradigm used here. In this study, normal subjects first learned to associate numbers to colors. In a second phase the correct associations were changed, and associations were tested under different degrees of arousal (varied by changing degree of induced muscular tension). At the time that arousal was varied, the responses which had previously been correct (regressive responses) were the strongest competing responses. The arousal paradigm implies that when arousal increases the strength of competing responses, the multiplicative effect of arousal should favor the strongest competing response (Levin's regression response) to the point where this response is restricted by the response-strength ceiling. Arousal increased beyond this level would tend to increase the proportion of other, initially weaker, competing responses. In the Levin study when proportion of regression errors to other errors was related to arousal level, the inverted U relationship implied by the present account was found. The proportion of errors which were regressive responses increased as muscle tension increased from zero to low to medium tension, and then decreased under high tension.

Another implication from the theoretical account used in the present paper is that regression is more likely to occur if both presently dominant and older responses have high response strength. Clinical literature and research reports (using animal subjects) have concentrated on only part of this implication, being concerned with the strength of the older or regressive response as a determiner of regression. Most authors have agreed that there is a positive relation between strength of the former response and likelihood of regression (e.g., Fenichel, 1945; Freud, 1949; Martin, 1940).

Thus there is some support for the hypothesis that a form of regression, the destructured mixture of mature, generalized, and previously appropriate behaviors which seems to occur in schizophrenia, may be the result of the same arousal-related process which determines other aspects of a schizophrenic syndrome.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

BARKER, R., DEMBO, T., & LEWIN, K. Frustration and regression: An experiment with young children. *University of Iowa Studies in Child Welfare,* 1941, **18,** No. 1.

BERNHAUT, M., GELLHORN, E., & RASMUSSEN, A. T. Experimental contributions to problem of consciousness. *Journal of Neurophysiology,* 1953, **16,** 21–35.

BLAUFARB, H. A demonstration of verbal abstracting ability in chronic schizophrenics under enriched stimulus and instructional conditions. *Journal of Consulting Psychology,* 1962, **26,** 471–475.

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: International Universities Press, 1950.

BROEN, W. E., JR., & STORMS, L. H. A reaction potential ceiling and response decrements in complex situations. *Psychological Review,* 1961, **68,** 405–415.

BROEN, W. E., JR., & STORMS, L. H. The differential effect of induced muscular tension (drive) on discrimination in schizophrenics and normals. *Journal of Abnormal and Social Psychology,* 1964, **68,** 349–353.

BROEN, W. E., JR., STORMS, L. H., & GOLDBERG, D. H. Decreased discrimination as a function of increased drive. *Journal of Abnormal and Social Psychology,* 1963, **67,** 266–273.

BROEN, W. E., JR., STORMS, L. H., & SCHENCK, H. U., JR. Inappropriate behavior as a function of the energizing effect of drive. *Journal of Personality,* 1961, **29,** 489–498.

BURSTEIN, A. G. Some verbal aspects of primary process thought in schizophrenia. *Journal of Abnormal and Social Psychology,* 1961, **62,** 155–157.

CAMERON, N. Deterioration and regression in schizophrenic thinking. *Journal of Abnormal and Social Psychology,* 1939, **34,** 265–269.

CAMERON, N. Experimental analysis of schizophrenic thinking. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley, Los Angeles: University of California Press, 1944. Pp. 50–63.

CAMERON, N. Perceptual organization and behavior pathology. In R. R. Blake & G. V. Ramsey (Eds.), *Perception—an approach to personality.* New York: Ronald, 1951. Pp. 283–306.

CARLTON, P. L. Cholinergic mechanisms in the control of behavior by the brain. *Psychological Review,* 1963, **70,** 19–39.

CAVANAUGH, D. Improvement in the performance of schizophrenics on concept formation tasks as a function of motivational change. *Journal of Abnormal and Social Psychology,* 1958, **57,** 8–12.

CHAPMAN, L. J. Distractibility in the conceptual performance of schizophrenics. *Journal of Abnormal and Social Psychology,* 1956, **53,** 286–291.

CHAPMAN, L. J. Intrusion of associative responses into schizophrenic conceptual performance. *Journal of Abnormal and Social Psychology,* 1958, **56,** 374–379.

CHAPMAN, L. J. A reinterpretation of some pathological disturbances in conceptual breadth. *Journal of Abnormal and Social Psychology,* 1961, **62,** 514–519.

CHAPMAN, L. J., BURSTEIN, A. G., DAY, D., & VERDONE, P. Regression and disorders of thought. *Journal of Abnormal and Social Psychology,* 1961, **63,** 540–545.

CHAPMAN, L. J., & CHAPMAN, J. P. Interpretation of words in schizophrenia. *Journal of Personality and Social Psychology,* 1965, **1,** 135–146.

CHAPMAN, L. J., & TAYLOR, J. A. Breadth of deviant concepts used by schizophrenics. *Journal of Abnormal and Social Psychology,* 1957, **54,** 118–123.

DEERING, G. Affective stimuli and disturbance of thought processes. *Journal of Consulting Psychology,* 1963, **27,** 338–343.

DEMILLE, R. Learning theory and schizophrenia: A comment. *Psychological Bulletin,* 1959, **56,** 313–314.

DISTLER, L. S., MAY, P. R. A., & TUMA, A. H. Anxiety and ego strength as predictors of response to treatment in schizophrenic patients. *Journal of Consulting Psychology,* 1964, **28,** 170–177.

DUNN, W. L. Visual discrimination of schizophrenic subjects as a function of stimulus meaning. *Journal of Personality,* 1954, **23,** 48–64.

EVARTS, E. V. A neurophysiologic theory of hallucinations. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 1–14.

FAIBISH, G. M. Schizophrenic response to words of multiple meaning. *Journal of Personality,* 1961, **29,** 414–427.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FEY, E. T. The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. *Journal of Consulting Psychology,* 1951, **15,** 311–319.

FISH, F. A neurophysiological theory of schizophrenia. *Journal of Mental Science,* 1961, **107,** 828–838.

FREEMAN, T. On the psychopathology of schizophrenia. *Journal of Mental Science,* 1960, **106,** 925–937.

FREUD, S. *An outline of psychoanalysis.* New York: Norton, 1949.

GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *Journal of Personality,* 1952, **20,** 253–276.

GELLHORN, E. Motion and emotion: The role of proprioception in the physiology and pathology of the emotions. *Psychological Review,* 1964, **71,** 457–472.

GOTTESMAN, L., & CHAPMAN, L. J. Syllogistic reasoning errors in schizophrenia. *Journal of Consulting Psychology,* 1960, **24,** 250–255.

HARRIS, A. Sensory deprivation and schizophrenia. *Journal of Mental Science,* 1959, **105,** 235–237.

HILL, W. F. Comments on Taylor's "Drive theory and manifest anxiety." *Psychological Bulletin,* 1957, **54,** 490–493.

KAUSLER, D. H., LAIR, C. V., & MATSUMOTO, R. Interference transfer paradigms and the performance of schizophrenics and controls. *Journal of Abnormal and Social Psychology,* 1964, **69,** 584–587.

KLEEMEIER, R. W. Fixation and regression in the rat. *Psychological Monographs,* 1942, **54**(4, Whole No. 246).

KNOPF, I. J., & FAGER, R. E. Differences in gradients of stimulus generalization as a function of psychiatric disorder. *Journal of Abnormal and Social Psychology,* 1959, **59,** 73–76.

KRAEPELIN, E. *Dementia praecox and paraphrenia.* (Trans by R. M. Barclay) (Ed. by G. M. Robertson) Edinburgh: Livingstone, 1919.

LEVIN, I. P. Induced muscle tension and response shift in paired-associate learning. Unpublished doctoral dissertation, University of California, Los Angeles, 1965.

LINDSLEY, O. R. Direct measurement and functional definition of vocal hallucinatory symptoms. *Journal of Nervous and Mental Disease,* 1963, **136,** 293–297.

LYNN, R. Russian theory and research on schizophrenia. *Psychological Bulletin,* 1963, **60,** 486–498.

MALITZ, S., WILKENS, B., & ESECOVER, H. A comparison of drug-induced hallucinations with those seen in spontaneously occurring psychoses. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 50–63.

MALMO, R. B. Activation: A neuropsychological dimension. *Psychological Review,* 1959, **66,** 367–386.

MARTIN, R. F. "Native" traits and regression in the rat. *Journal of Comparative Psychology,* 1940, **30,** 1–16.

MEDNICK, S. A. A learning approach to research in schizophrenia. *Psychological Bulletin,* 1958, **55,** 316–327.

MEEHL, P. E. Schizotaxia, schizotypy, schizophrenia. *American Psychologist,* 1962, **17,** 827–838.

NOYES, A. P., & KOLB, L. C. *Modern clinical psychiatry.* Philadelphia: Saunders, 1958.

PAYNE, R. W. Cognitive abnormalities. In H. J. Eysenck (Ed.), *Handbook of abnormal psychology.* New York: Basic Books, 1961. Pp. 193–261.

PAYNE, R. W., CAIRD, W. K., & LAVERTY, S. G. Overinclusive thinking and delusions in schizophrenic patients. *Journal of Abnormal and Social Psychology,* 1964, **68,** 562–566.

PEASTREL, A. L. Studies in efficiency: Semantic generalization in schizophrenia. *Journal of Abnormal and Social Psychology,* 1964, **69,** 444–449.

PINNEO, L. R. The effects of induced muscle tension during tracking on level of activation and on performance. *Journal of Experimental Psychology,* 1961, **62,** 523–531.

RECHTSCHAFFEN, A., SCHULSINGER, F., & MEDNICK, S. A. Schizophrenia and physiological indices of dreaming. *Archives of General Psychiatry,* 1964, **10,** 89–93.

RISS, E. Are hallucinations illusions? An experimental study of nonveridical perception. *Journal of Psychology,* 1959, **48,** 367–373.

RODNICK, E. H. The effect of metrazol shock upon habit systems. *Journal of Abnormal and Social Psychology,* 1942, **37,** 560–565.

SADLER, W. S. *Practice of psychiatry.* St. Louis: Mosby, 1953.

SCHNORE, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *Journal of Experimental Psychology*, 1959, **58**, 117–128.

SILVERMAN, J. Scanning-control mechanism and "cognitive filtering" in paranoid and nonparanoid schizophrenia. *Journal of Consulting Psychology*, 1964, **28**, 385–393.

SOMMER, R., DEWAR, R., & OSMOND, H. Is there a schizophrenic language? *Archives of General Psychiatry*, 1960, **3**, 665–673.

SPENCE, J. T., & LAIR, C. V. Associative interference in the verbal learning performance of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1964, **68**, 204–209.

SPENCE, J. T., & LAIR, C. V. Associative interference in paired-associate learning of remitted and nonremitted schizophrenics. *Journal of Abnormal Psychology*, 1965, **70**, 119–122.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale University Press, 1956.

SPENCE, K. W. Behavior theory and selective learning. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: University of Nebraska Press, 1958. Pp. 73–107. (a)

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *American Psychologist*, 1958, **13**, 131–141. (b)

STILSON, D. W., & KOPELL, B. S. The recognition of visual signals in the presence of visual noise by psychiatric patients. *Journal of Nervous and Mental Disease*, 1964, **139**, 209–221.

STORMS, L. H., & BROEN, W. E., JR. Verbal associative stability and appropriateness in schizophrenics, neurotics, and normals as a function of time pressure. *American Psychologist*, 1964, **19**, 460. (Abstract)

TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety to level of performance in serial learning. *Journal of Experimental Psychology*, 1952, **44**, 61–64.

TUTKO, T. A., & SPENCE, J. T. The performance of process and reactive schizophrenics and brain injured subjects on a conceptual task. *Journal of Abnormal and Social Psychology*, 1962, **65**, 387–394.

VENABLES, P. H. The effect of auditory and visual stimulation on skin potential responses of schizophrenics. *Brain*, 1960, **83**, 77–92.

WECKOWICZ, T. E., & BLEWETT, D. B. Size constancy and abstract thinking in schizophrenic patients. *Journal of Mental Science*, 1959, **105**, 909–934.

WEST, L. J. A general theory of hallucinations and dreams. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 275–291.

ZASLOW, R. W. A new approach to the problem of conceptual thinking in schizophrenia. *Journal of Consulting Psychology*, 1950, **14**, 335–339.

ZUCKERMAN, M., ALBRIGHT, R. J., MARKS, C. S., & MILLER, G. L. Stress and hallucinatory effects of perceptual isolation and confinement. *Psychological Monographs*, 1962, **76**(30, Whole No. 549).

(Received March 10, 1965)

Psychological Review
1966, Vol. 73, No. 4, 280–294

CUTANEOUS COLOR SENSITIVITY: EXPLANATION AND DEMONSTRATION

W. L. MAKOUS[1]

IBM Research Center, Yorktown Heights, New York

Past reports that humans can, in complete darkness, sense with their fingers the colors objects would have if illuminated, have understandably been received with skepticism. A previously proposed hypothesis, based on differential rate of absorption of infrared radiation by different layers of the skin, is inconsistent with the thermodynamics of the situation. Quantitative analysis of the system consisting of a room-temperature surface juxtaposed to the higher temperature skin, however, leads to the conclusion that large differences in emissivity of different room-temperature surfaces almost certainly could be detected by the associated effects on skin temperature. Easily repeatable demonstrations show that this is true, and a few simple temperature measurements confirm the theoretical identification of the mechanism.

Occasionally over the years popular accounts and clinical reports have appeared that describe a rather incredible capacity of some individuals to sense with their fingers the properties of objects that normally only the eyes can detect. Recent investigations, beginning with those of Goldberg (1963), Youtz (1963, 1964), and Nyuberg (1964), have increased the creditability of such reports to the point where they must be investigated to explain their apparent incompatibility with the present concepts of sensory function.

Some of these reports are probably explained by the inadequacy of blindfolds for preventing use of the eyes (see Gardner, 1966). Among these is Nyuberg's (1964) report on an individual who identified printed characters and demonstrated trichromatic color sensitivity while blindfolded. In all respects, the subject's performance was remarkably like what one would expect if it were mediated by images reaching the retina through the pinhole apertures that typically exist at the edges of blindfolds (Gardner, 1966). Goldberg (1963) reported earlier tests on the same individual under conditions that would seem to make peeking more difficult, but his report generally lacks signs of the rigor required in an investigation of unlikely phenomena. In such cases, where the results seem very unlikely, and especially in cases where a small number of individuals claim an extraordinary ability, rather extensive precautions against trickery may be required as is advised by Gardner (1966). In all cases in which light is intended as a cutaneous stimulus (e.g., in Leont'ev's work described by Pick, 1964, p. 22), interpretation of the results must take into account the thermal effects of light on skin; "heat filters" do not eliminate these effects, for no filter can pass light and not pass energy which can be converted to heat upon absorption.

Youtz's (1963, 1964, 1965) results are not so easily explained as those mentioned above. He tested an exceptional individual who was able to identify the color of paper, plastics, and cloth, which were in the complete darkness of the interior of a light-tight box. (Color is used here in the technical sense, encompassing brightness as well as hue.) The subject was required to wear a blindfold and to rest her forehead against the box more than a foot above two holes in the side of the box, to which were attached thick, black sleeves with elastic wrist bands, through which the subject was required to thrust her hands in order to make the identifications. Four different series of experiments were conducted in this manner. The subject was correct 85–95% of the time in the first series, much less often but still significantly above chance in the second series, and at chance level in the third and fourth series.

---

[1] The many helpful comments of colleagues and the technical assistance of Jean Kirby are gratefully acknowledged. The author's present addresses is Department of Psychology, University of Washington, Seattle, Washington.

Youtz (1964, 1965) also tested 133 Barnard College freshmen under the following conditions: To each he attached a blindfold and also a "bib" stretching from around the subject's neck to the top of the frame holding the test objects, thus completely hiding the test objects from the subject. He reports that 10% of the subjects were able to perform under these conditions consistently (not Type I errors) above chance. To explain the performance of these subjects, Youtz has advanced what he calls the "wavelength-temperature discrimination hypothesis" (1964, 1965).

Unfortunately, Youtz has not yet published a full account of this work so that his procedure and analysis can be carefully studied. Thus one could attribute the gradual loss of ability of his exceptional subject to learning on the part of the experimenter, and the probability statistics in his study of Barnard students are not reported in enough detail to make the data so far reported thoroughly convincing. Yet, although the wavelength-temperature discrimination hypothesis is almost certainly incorrect, as shown below, the conclusion that humans can perform discriminations of the type described by Youtz almost certainly is correct.

The purposes of this paper, then, are:

1. To point out the flaws in the wavelength-temperature discrimination hypothesis.
2. To show that we should not be surprised if people can discriminate objects in the dark, which appear to differ only in color when illuminated; however, certain thermodynamic considerations and established sensory capacities would lead us to be surprised if people could not make such discriminations.
3. To describe a simple demonstration whereby the skeptical reader can prove to himself his ability to make such discriminations.

WAVELENGTH-TEMPERATURE DISCRIMINATION HYPOTHESIS

The inevitable consequence of absorption of radiant energy, whether in the visible spectrum or not, is an increase in temperature of the absorbing entity. We know that the human skin is exquisitely sensitive to temperature changes: Under some conditions a human subject can detect a cutaneous temperature change of .001°C sec^{-1} (Hardy & Oppel, 1937; Hendler & Hardy, 1960; Hendler, Hardy, & Murgatroyd, 1961; Hensel, 1950). Thus it is of no special interest that a subject can detect the differences caused by illumination of the skin by different light colors unless the difference in energy absorbed is smaller than is likely to be detected by the resulting effects on skin temperature, a capacity that has not yet been convincingly demonstrated.

Discrimination of objects that are neither illuminated nor self-luminous poses a different problem, however, for the way in which the object serves as an energy source may not be obvious. To account for discrimination of such objects, Youtz (1964, 1965) has advanced the following hypothesis.

The human skin, a nearly ideal black-body radiator under ordinary conditions (Barnes, 1963; Hardy, 1934b; Hardy & Muschenheim, 1934), radiates energy within the dark, light-tight box containing the test objects. The energy radiated by the skin is absorbed by the test objects and reemitted at wavelengths determined by their respective compositions, particularly by any dyes or pigments present. The rays reemitted from the test objects penetrate the skin to varying depths before being absorbed, depending upon their wavelengths (Hardy & Muschenheim, 1934, 1936; Oppel & Hardy, 1937a, 1937b). Localized heating then occurs at the depths at which the rays are absorbed, and the subject hypothetically discriminates the relative depths at which the heating occurs. Necessary for this hypothesis, then, are: (*a*) detectable radiation by the test objects in wavelengths which penetrate skin to different depths, and (*b*) the human ability to discriminate slight differences in temperature, or differences in temperature change, at different depths within the skin.

This hypothetical explanation is plausible enough. The lack of independent evidence that humans can make the necessary *b* discrimination is only a minor drawback. It is shown below, however, that the hypothesis fails to satisfy condition *a;* for, under the conditions of Youtz's experiments, practically all of the energy is radiated by the test objects in wavelengths that are absorbed too close to the skin surface to provide depth cues.

Thermodynamic Considerations

The foregoing conclusion follows from certain thermodynamic considerations that hold for all but very special conditions such as are artificially arranged within the laboratory or are inferred to exist in exotic places like the interiors of stars. Since the hypothesis concerns only radiant heat exchanges, discussion of the effects of conduction and convection will be deferred until later. The principal point here is that a substance of a given temperature, such as the subject's skin or one of the test objects, cannot radiate more energy at any wavelength than does an ideal radiator at the same (or higher) temperature, unless it receives energy from a higher energy source. If the substance in the lower energy state (the test object) receives energy from the substance in the higher energy state (the skin) only through radiation, to which the present case is limited, then the energy absorbed cannot be reemitted (fluoresced) at a shorter wavelength than the wavelength at which it was originally absorbed (Stokes' law). This means that so long as the test objects are cooler than the skin, they always return less energy to the skin in wavelengths shorter than any particular wavelength than the skin emits in those wavelengths; this, then, places an upper limit on the amount by which test objects can differ from one another in their radiant exchanges with the skin. These limits will now be taken up quantitatively.

The proportion of incident radiation absorbed rather than reflected by an opaque object at a given wavelength is expressed by its spectral emissivity, ϵ_λ. A perfect reflector reflects all incident radiation and therefore absorbs none ($\epsilon_\lambda = 0$): An ideal black-body radiator absorbs all incident radiation ($\epsilon_{\lambda_b} = 1.0$), but it emits energy according to Planck's radiation formula:

$$H_{\lambda_b} = \frac{c_1\lambda^{-5}}{e^{c_2/\lambda T_b} - 1},$$

where H_λ = spectral radiant flux density in w cm^{-1} mμ^{-1}, λ = wavelength in μ, T = temperature in K°, $c_1 = 37.403$ w cm^{-1} mμ^{-1}, $c_2 = 14387.9\ \mu°$ K.

The energy radiated by the skin at a given wavelength, H_{λ_s}, equals the product of its spectral emissivity, ϵ_{λ_s}, and the energy radiated at that wavelength and temperature by a black body, H_{λ_b}:

$$H_{\lambda_s} = \epsilon_{\lambda_s} H_{\lambda_b} = \frac{\epsilon_{\lambda_s} c_1\lambda^{-5}}{e^{c_2/\lambda T_s} - 1}.$$

Then the maximum amount of energy of wavelength λ that could be returned to the skin by a juxtaposed test object, if it were a perfect reflector, is the total amount emitted by the skin, H_{λ_s}. (To simplify geometric considerations, the distance of the test object from the skin is assumed to be small relative to its length and width.) And the least amount of energy of wavelength λ that could be returned to the skin by a juxtaposed test object, if it were an ideal radiator, is H_{λ_b}, of which $\epsilon_{\lambda_s} H_{\lambda_b}$ would be absorbed by the skin.

The net energy loss of the skin over the interval, $\lambda = a$ to $\lambda = z$, is the difference between the energy radiated and the energy absorbed over that interval. The least that loss can be, when the skin is juxtaposed to a perfect reflector, is zero; and the greatest it can be, when juxtaposed to an ideal radiator, is:

$$\Delta H_{\max} = -\int_a^z H_{\lambda_s} d\lambda + \int_a^z \epsilon_{\lambda_s} H_{\lambda_b} d\lambda$$
$$= -\int_a^z \frac{\epsilon_{\lambda_s} c_1 \lambda^{-5} d\lambda}{e^{c_2/\lambda T_s} - 1} + \int_a^z \frac{\epsilon_{\lambda_s} c_1 \lambda^{-5} d\lambda}{e^{c_2/\lambda T_b} - 1}.$$

Then the interval, $\Delta H_{\max}$, is the maximum range over which the test objects can vary in their effects on the skin. From this equation it is clear that this range depends upon the spectral emissivity of the skin, ϵ_{λ_s}; upon the temperature of the skin, T_s; upon the temperature of the test objects, T_b; and upon the spectral band considered, $a \leq \lambda \leq z$.

Computations

Radiation in different wavelengths can be classified according to whether skin is relatively transparent to it, relatively opaque to it, or of intermediate density with respect to it. These three categories partition the relevant part of the spectrum into three regions (Hardy, Hammel, & Murgatroyd, 1956; Hardy & Muschenheim, 1934, 1936; Hendler & Hardy, 1960; Hendler et al., 1961; Oppel & Hardy, 1937a, 1937b): Wavelengths that are, respectively, shorter than .8 μ, longer than 3 μ, and between .8 μ and 3 μ. For expository purposes, and because Youtz has implicated these three regions in his hypothesis, each will be considered separately.

It is clear at the outset that practically no energy is emitted by the skin in the wavelengths shorter than .8 μ, that is,

$$\Delta H_{\max} = \int_0^{.8} \epsilon_{\lambda_s} H_{\lambda_s} d\lambda \simeq 0,$$

for the eyes are so sensitive to these

wavelengths that we should otherwise be able to see one another in the dark. To bias the result in favor of the wavelength-temperature discrimination hypothesis, let ϵ_{λ_s} and T_s assume their maximum respective values of 1.0 and 98.6°F and let,

$$\int_0^{.8} \epsilon_{\lambda_s} H_{\lambda_b} d\lambda = 0.$$

Then,

$$\Delta H_{\max} = -\int_0^{.8} H_{\lambda_s} d\lambda$$
$$= -1.13(10^{-22}) \text{ w cm}^{-2}.$$

(The negative sign indicates that net energy flux is outward from the skin.) Since the energies of photons in these wavelengths exceed 2.4×10^{-19} j/photon, the ratio,

$$\frac{-1.13(10^{-22}) \text{ w cm}^{-2}}{2.4(10^{-19}) \; j/\text{photon}}$$
$$= 4.7(10^{-4}) \text{ photon sec}^{-1} \text{ cm}^{-2}$$
$$= 1.7 \text{ photon hr}^{-1} \text{ cm}^{-2},$$

shows that a square centimeter of skin at body temperature loses, on the average, no more than two photons in these wavelengths every hour.

Consider now radiation in the wavelengths between .8 μ and 3 μ, which penetrates to intermediate depths below the skin surface. The spectral emissivity of the human skin in these wavelengths is reported to be between .75 and .95 (Barnes, 1963; Hardy, 1934b; Hardy & Muschenheim, 1934; Oppel & Hardy, 1937a), but again to bias the result in favor of the wavelength-temperature discrimination hypothesis, let ϵ_{λ_s} and T_s assume their respective maximum values of 1.0 and 98.6°F, and let T_b be a room temperature of 72°F. Then

$$\Delta H_{\max}$$
$$= -\int_0^{3} \frac{c_1\lambda^{-5} d\lambda}{e^{c_2/\lambda T_s} - 1} + \int_0^{3} \frac{c_1\lambda^{-5} d\lambda}{e^{c_2/\lambda T_b} - 1}$$
$$= -7.5(10^{-6}) + 3.3(10^{-6})$$
$$= -4.2(10^{-6}) \text{ w cm}^{-2}$$
$$= -10^{-6} \text{ g-cal sec}^{-1} \text{ cm}^{-2}.$$

The threshold radiant flux density for large skin areas, H_{thresh}, depends upon a number of variables, but in this situation it lies between 1.5×10^{-4} and 7.0×10^{-4} g-cal sec^{-1} cm^{-2} (Hardy & Oppel, 1937; Hendler & Hardy, 1960; Hendler et al., 1961; Lele, Weddell, & Williams, 1954). This is greater by two orders of magnitude than $\Delta H_{\max}$, the maximum possible energy exchange that could take place between the skin and a room-temperature object via wavelengths shorter than 3 μ. One might argue that less energy is required to reach threshold at these wavelengths than at the longer wavelengths used in arriving at the above value of H_{thresh}, but just the opposite appears to be the case (Oppel & Hardy, 1937a).

If, nevertheless, there is some error in the conclusion that the signal afforded by 10^{-6} g-cal sec^{-1} cm^{-2} is too weak to be detected, then under the conditions here this signal would almost certainly be lost in the noise associated with the much larger energy exchanges occurring via the longer wavelengths ($\lambda > 3\ \mu$). These exchanges,

$$\Delta H_{\max}$$
$$= -\int_3^{\infty} H_{\lambda_s} d\lambda + \int_3^{\infty} H_{\lambda_b} d\lambda$$
$$= 9.3(10^{-3}) \text{ w cm}^{-2}$$
$$= 2.2(10^{-3}) \text{ g-cal sec}^{-1} \text{ cm}^{-2},$$

are more than 2000 times those occurring in the shorter wavelengths just discussed. Computations show that variations of ±.013°F in T_b or T_s, or variations of ±.001 in emissivity either of the skin or of the test objects, would cause variations in the value of $\Delta H_{\max}$ for the wavelengths longer than $3\,\mu$, as large as the entire value of $\Delta H_{\max}$ for the wavelengths shorter than $3\,\mu$.

Then no detectable radiant exchange occurs via wavelengths shorter than $3\,\mu$, and if depth of penetration provides the subjects with any information, it must occur via wavelengths longer than $3\,\mu$. But all wavelengths longer than $3\,\mu$ are absorbed at approximately the same depth (Hardy & Muschenheim, 1934). Furthermore, 90% of this radiation is absorbed within the outermost $30\,\mu$ or so of skin (Hardy & Muschenheim, 1936), which is keratinized tissue devoid of nerve endings and receptors (Ranson & Clark, 1953). Little more than 1% of incident radiation in these wavelengths penetrates even so far as the most superficial skin layers suspected of harboring thermal receptors, more than $100\,\mu$ below the surface (Hendler et al., 1961; Zotterman, 1959). That radiation in these wavelengths does not penetrate any significant distance is supported by the finding (Hendler et al., 1961) that blackening the skin surface, which insures complete absorption of incident radiation at the very surface of the skin, has no effect upon thresholds to such radiation.

To recapitulate, then, a necessary condition for the temperature-wavelength discrimination hypothesis is that the test objects emit detectable radiation at wavelengths that penetrate to different depths within the skin. But, as just demonstrated, practically all of the energy is exchanged via wavelengths that are absorbed at a uniform depth, and that depth is very close to the surface but distant from receptors. Thus, the necessary condition is not satisfied, and the wavelength-temperature discrimination hypothesis cannot hold.

Mechanism

Despite the flaws of the wavelength-temperature discrimination hypothesis, it has led to an understanding of the situation it was intended to explain. The total radiant heat exchange between the skin at body temperature and a black body at room temperature, 2.2×10^{-3} g-cal sec^{-1} cm^{-2}, is anywhere from 3–15 times threshold, and thus it is certainly large enough to be detected. Objects having sufficiently different spectral emissivities in the infrared, when held close to the skin, should produce detectably different rates of skin cooling, so long as a sufficient temperature difference is maintained between them and the skin. That subjects use this temperature cue as the basis of their discriminations is supported by some of their subjective reports (Youtz, 1964).

One can, in fact, compute the differences in emissivity that should be detectable on this basis with any given combination of temperatures. Let real substances, having spectral emissivity, ϵ_{λ_t}, replace the ideal substances in the equation for $\Delta H_{\max}$, and let $\epsilon_{\lambda_s} = 1.0$, $a = 0$, and $z = \infty$, then the net energy loss by the skin is the difference between the total energy emitted by the skin and the sum of energy emitted and the energy reflected by the substance, all of which is absorbed by the skin (because $\epsilon_{\lambda_s} = 1.0$). Thus,

$$\begin{aligned}\Delta H = -\int_0^\infty H_{\lambda_s} d\lambda &+ \int_0^\infty \epsilon_{\lambda_t} H_{\lambda_t} d\lambda \\ &+ \int_0^\infty (1 - \epsilon_{\lambda_t}) H_{\lambda_s} d\lambda \\ &= \int_0^\infty \epsilon_{\lambda_t} (H_{\lambda_t} - H_{\lambda_s}) d\lambda.\end{aligned}$$

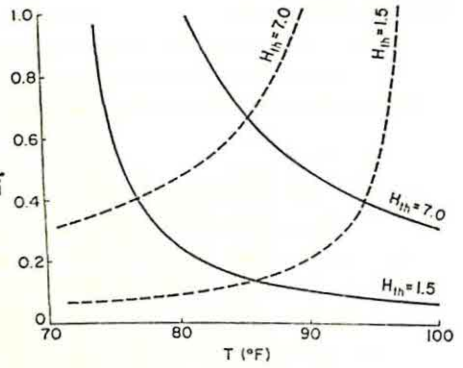

FIG. 1. Minimum detectable emissivity difference as a function of temperature. (The independent variable for the solid curves is skin temperature, and $T_t = 72$°F; the independent variable for the broken curves is temperature of the test object, and $T_s = 98.6$°F.)

The value of ΔH obviously depends on the spectral absorption curve of the particular test object under consideration. Not the most interesting, but nevertheless an instructive case results if $d\epsilon_{\lambda_t}/d\lambda = 0$; that is, if ϵ_{λ_t} is constant. This case is instructive because it provides a basis for estimating the approximate mean difference between spectral emissivities, weighted by Planck's radiation formula, that should be detectable under various circumstances. Let the difference in emissivity of the two test objects be written $\Delta\epsilon_t$. Then threshold,

$$H_{\text{thresh}} = \int_0^\infty \Delta\epsilon_t(H_{\lambda_t} - H_{\lambda_s})d\lambda$$
$$= \Delta\epsilon_t \int_0^\infty (H_{\lambda_t} - H_{\lambda_s})d\lambda.$$

Integration results in the Stefan-Boltzmann law,

$$H_{\text{thresh}} = \Delta\epsilon_t\sigma(T_t^4 - T_s^4),$$
$$\sigma = 5.67(10^{-12}) \text{ w cm}^{-2} \text{ K}^{-4}.$$

Then:

$$\Delta\epsilon_t = H_{\text{thresh}}/[\sigma(T_t^4 - T_s^4)].$$

The solid lines in Figure 1 show the effects of skin temperature, T_s, on the likely upper and lower limits of $\Delta\epsilon_t$, corresponding to the upper and lower estimates of threshold, H_{thresh}, $7.0(10^{-4})$ and $1.5(10^{-4})$ g-cal sec^{-1} cm^{-2}, respectively. Here, temperature of the test objects, T_t, is 72°F. The upper limiting curve shows that skin at body temperature should mediate discriminations of objects differing by .32 in their infrared emissivity. When skin temperature drops to 83°F, the emissivity of different objects may have to differ by .8 or more to be discriminated. This is consistent with the reports that subjects cannot perform these discriminations when their fingers are cold (Youtz, 1964). On the other hand, if the threshold is closer to the lower limit, very fine discriminations of objects differing in emissivity only by about .10 should be possible for a large part of the normal range of skin temperature (down to 90°F), and even down to 76°F, only 4°F above the temperature of the test objects, differences of .50 in emissivity would be detectable.

The broken lines in Figure 1 show the analogous curves when temperature of the test objects, T_t, varies, and skin temperature, T_s, is held at 98.6°F. It is obvious from these curves that the discrimination would be more difficult, for example, on a warm summer day in a room lacking air conditioning. It is even possible that this accounts for some of the difficulty experienced by Youtz's exceptional subject in the last series of experiments, done in the summer of 1964.

The discussion so far has concerned only discrimination of objects, but to a limited extent identification of their color might also be possible. Red objects have low spectral emissivities in the longer wavelengths of the visible spectrum, but blue objects have high emissivities in these wavelengths. Consequently, red substances are more likely to have low emissivities in the nearby infrared than are blue substances. Yellows would occupy an intermediate position in this respect. Examples of such correlations can be found in tables on transmission of colored glasses (Hodgman, Weast, & Selby, 1960). Thus, if the subject were *set* to interpret differences in induced skin temperature in terms of hue, and if reference stimuli were available, limited but statistically reliable identifications should be possible, mostly those at the ends of the visual spectrum, in the reds and blues. This is precisely the nature of the identifications reported (Youtz, 1964). That is, the subjects reportedly made limited but statistically reliable identifications of reds and blues only. It is also interesting in this connection that people usually consider red a "warm" color, and blue a "cool" color, which is in the direction consistent with this analysis.

DEMONSTRATION

The effect discussed above is great enough that one can make the simple preparations necessary and test it for himself in 1 or 2 hours.[2] The idea behind this demonstration is to maximize the effects described above in order to establish the effectiveness of the principle involved. After the effect is revealed under the optimum conditions, determining the limits of the effect and determining such things as the variability between individuals are straightforward, research questions.

Materials

Two surfaces are required, one approaching an ideal radiator, and the other approaching a perfect reflector. Polished, unoxidized metals, having emissivities of well under .10 (Hodgman et al., 1960) can be used as the reflector. A first-surface mirror is also suitable, but the ordinary second-surface mirror is not, because infrared rays are absorbed by the glass in the pathway to and from the reflecting surface. A reflector in the form of a fairly thick ($\frac{1}{4}$ inch or so), 6-inch square plate is especially convenient, as the radiator can be formed simply by painting half of it flat black, which gives that half an emissivity of .90–.95 (Hodgman et al., 1960). Next a frame is required that is thicker than the square test plate by about $\frac{1}{2}$ inch; it should fit loosely around the test plate, so that the plate can easily be inserted into the frame and removed from it. The last requirement is a table with an opaque screen at one end, under which one can thrust his arms without being able to see the table. An opaque laboratory apron (some are available with optical densities exceeding 10), tied about the neck and thrown back over the head, though uncomfortable, can serve in place of the opaque screen. A completely dark room (or less than 10^{-7} footcandles illuminance) is an added luxury. If there is reason to suspect a subject of trickery, an aluminum box of the type Gardner (1966) recommends can be used.

Procedure

The first step is to place the plate inside the frame without the subject's knowing on which side the painted half is located. If the plate is sufficiently uniform, one can both serve as subject and position the plate after first spinning it on the table, being careful not to receive tactile cues by touching the half-painted top surface. It is better, of course, to have someone else position the plate, perhaps using a set of random numbers. Chances of suggestion, or even of the possibility of telepathy, can be reduced by having the other person, who does not know its orientation until he checks it after the subject has made his choice, spin the square. A mechanical means of randomizing the plate orientation that provides no extraneous cues would be best of all, but the above techniques are adequate to demonstrate the point here.

The subject makes his choice by resting the ball of his hand on one side of the frame and the tips of his fingers on the other side; this permits him to hold his hand about $\frac{1}{2}$ inch

---

[2] In fact, it would make a good experiment or demonstration for an elementary psychology laboratory.

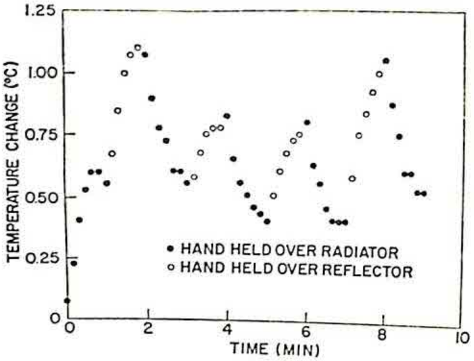

FIG. 2. Skin temperature changes as a function of time juxtaposed alternately to a reflecting surface and to an absorbing surface. (See text.)

above the test plate without touching it. Although it is not necessary, a simple detector can be attached so that a trial can be stopped before the subject has made his choice if he happens to touch the test plate. This is done simply by putting 1.5 volts across a bridge and monitoring the bridge balance with an oscilloscope. The skin and the test plate are wired so that any momentary contact between the skin and either half of the test plate shorts a very large resister in the bridge, throwing it off balance. As the subject moves his hand back and forth, from side to side, leaving it in place a few seconds each time, he notices a feeling of warmth on one side (some subjects feel a coolness, or "draft" on the other side); if he identifies the warm side as the unpainted half (or the side where he notices the draft as the painted half), he will be correct about 90% of the time. The discrimination is slightly easier if the number of plate orientations is limited to two, but the probability of explanations based on chance rapidly becomes infinitesimal when four positions are permitted.

TYPICAL RESULTS

All of the options mentioned above, except for mechanical orientation, have been tried in 8 sets of 10 trials, using six subjects, one of whom was totally blind. Three of the subjects were tested while wearing an aluminum box of the type described by Gardner (1966). Each of the 8 sets individually attained statistical significance (binomial test, .05 level); the lowest score was 8 correct out of 10 trials, but because the plate had been given four possible orientations in that particular set, it was still significant (.001 level).

Temperature Measurements

The skin temperature of two subjects was measured with a thermocouple embedded in the palms of their hands, as their hands moved back and forth above the two halves of a 6-inch square first-surface mirror, half painted flat black. Precise measurement of skin temperature by this means is difficult (Barnes, 1963; Davis, 1961, 1963; Hardy, 1934a; Hardy & Stoll, 1954), but this technique is adequate for the purposes here.

Representative results are shown in Figure 2. The data points represent readings from a differential voltmeter, later converted to degrees centigrade from calibration curves. The ordinate shows changes in temperature rather than absolute temperature because most of the errors inherent in this technique appear as constant errors affecting measurements of absolute temperature, while relative changes in temperature are much more accurately reflected (Hardy, 1934a; Molner & Rosenbaum, 1961).

The figure shows that temperature varied over approximately .5°C, in synchrony with the movement of the hand from one half of the plate to the other half. The discrimination depends not so much upon the magnitude of temperature differences attained, however, as upon the rate of these changes (Hendler & Hardy, 1960; Hensel, 1950; Lele et al., 1954), as reflected in the slope of the curve. The maximum slopes typical of the transition from one temperature to another as the hand moved from above one surface to above the other were .017°C $\sec^{-1}$, well above threshold under these conditions (Hendler & Hardy, 1960; Hensel, 1950).

The fact that the temperature of the hand stayed well above the temperature it was at the beginning of the session (represented by the origin of the curve) shows that the equilibrium point of the hand is higher when its thermal exchanges with the ambient must pass through the test plate and its holder than when these exchanges are more direct, as one would expect.

OTHER VARIABLES

The effects of convection, conduction, and the cumulative effects of thermal exchange have been largely ignored in the discussion so far.

Convection

Convection is minimal in the demonstration described above, as the air is almost completely enclosed in a small compartment, and the thermal gradient produced within the air space between the skin and the test objects produces a density gradient that tends to maintain the status quo. Convection was prevented in Youtz's experiments by interposing between the skin and the test objects a layer of "transparent" plastic instead of air.

Conduction

The thermal conductivity of air is $.6 \times 10^{-4}$ g-cal cm $\sec^{-1}$ $\text{cm}^{-2}{}^\circ\text{C}^{-1}$ (Hodgman et al., 1960), and in the demonstration here, the hand is held about 1 centimeter from the surface of the test object. If the skin were 98.6°F and if the test object were 72°F, a difference of 15°C, then the air would conduct 9×10^{-4} g-cal $\sec^{-1}$ cm^{-2}. Air conduction of heat from the skin, then, is a linear function of the temperature difference between the skin and the test object, and it is in the order of magnitude of the absolute threshold.

The layer of paint covering half of the test plate, having a conductivity perhaps 1% that of the metal (Hodgman et al., 1960) may appear to be a complication in assessing the effects of thermal conduction to the test plate. One might even wonder whether the insulating effect of the paint might account for the discrimination rather than the radiant effects. The answer is that it does not, because the effect is in the wrong direction: The paint may retard, to some extent, the rate of thermal conduction when the hand is poised above the painted half of the test plate; the skin would therefore lose less heat when above that side and would feel warmer there. This is exactly opposite to the findings. The reason the paint is an ineffective insulator in this case is because it is so thin, and its conductivity is so much greater than that of the air that the heat conducted to it by the air is rapidly passed to the metal to which it adheres.

Thermal conduction in Youtz's experiment would have been primarily through a medium of plastic rather than air. The plastic, being a solid, like the paint, would have had a thermal conductivity more than 20 times that of air (Hodgman et al., 1960), and the plastic layer is not likely to have been as thick as 1 centimeter; consequently, thermal conduction would have had a much greater effect in his experiments than in the demonstration here. How much greater depends upon a number of things, among which is in the area of contact between the subjects' skin and the plastic covering. This analysis leads to the paradoxical conclusion that, the harder the subject pushed down on the test objects, the less he would have been able to discriminate between them.

Cumulative Effects

The specific heat, density, conductivity, and spatial dimensions of a body, along with its thermal interactions with other parts of its environs, determine the cumulative effects of thermal exchange with any other body. The test object in the demonstration is an effective heat sink because heat is rapidly distributed as it is received, and because its size, density, and specific heat permit it to absorb a lot of heat before its temperature rises greatly. Finally, because of its large surface area in comparison with its volume, the test plate rapidly loses heat to its surroundings by conduction and radiation, both of which are proportional to surface area.

On the other hand, a test object lacking these properties would behave like a series capacitor in an electrical circuit; it would pass only transients, or changes in heat flow. Thus, a subject would have only a short time in order to make his judgment before the object warmed up to a temperature close to that of his skin, where discrimination is poor (see Figure 1).

Moreover, because of the higher rate of heat exchange between the skin and the more emissive surface, the temperature of that surface would tend to rise more rapidly, and to attain higher values, than that of the more reflective surface. (Conversely, after the hand had been removed, the temperature of the more emissive surface would tend toward that of the ambient more rapidly.) As the temperature of the emissive surface rose an increasing distance above that of the reflective surface, the difference between the two corresponding rates of heat exchange with the skin would first diminish, next drop below threshold, then reverse, and finally rise above threshold in the opposite direction. Thus, if the temperature of the more emissive surface rose a sufficient distance above that of the more reflective surface, the skin would lose more heat to the more reflective surface than to the more emissive one, and the phenomenal effect would be reversed: the more emissive surface would feel warmer. The relevant variables were not sufficiently well controlled in the demonstration to warrant an analysis of the conditions under which this might occur, but the advantage of combining the test objects in a single piece of highly conductive material is not only that it provides a good heat sink, but also that it maintains the two surfaces at a more nearly equal temperature than would otherwise be possible without thermoregulatory equipment.

In the demonstration, the vascular system provides the ultimate heat source. The cutaneous thermal receptor mechanisms lying between the blood supply and the skin surface are complicated and controversial, and since the conclusions of this paper depend primarily upon stimulus-response relationships, these mechanisms will not be discussed. Readers are referred instead to three review papers for some of the different views receiving current attention (Nafe & Kenshalo, 1962; Weddell & Miller, 1962; Zotterman, 1959).

Optimum Conditions

The following list summarizes the conditions facilitating the discrimination of objects on the basis of their radiant properties: (*a*) The difference in spectral emissivities of the test objects, in the wavelengths from 4 μ to 20 μ, should be maximum; (*b*) the temperature difference between the test objects and the skin should be maximum, but the test objects themselves should be at the same temperature; (*c*) the test objects should be good heat sinks, but they should not differ in this respect; (*d*) conduction and convection of heat between the skin and the test objects should be minimum; (*e*) the medium between the skin and the test objects should pass radiation of $4\,\mu$ to $20\,\mu$ wavelengths; (*f*) the test objects should be slightly larger than the size of the hand, to take full advantage of "spatial summation" (Hardy & Oppel, 1937; Lele et al., 1954).

Suboptimum Conditions

A thermal basis for cutaneous discrimination between the test objects can be provided by differences in any of the six variables implicated in the preceding paragraph. Emissivity, however, is the most fundamental and perhaps the least obvious variable. Differences in more than one variable confound the effects of these variables in complicated ways that can produce opposite results under conditions that differ only subtly (e.g., see the section on Cumulative Effects).

Illumination of the test objects can also complicate the results. In general, illumination hinders the discrimination, both by augmenting the temperature rise of the test objects toward that of the skin, where discrimination is difficult (see Figure 1), and by increasing the temperature difference between the test objects (for discussion, see section on Cumulative Effects). Other complications follow from the different emission spectra of the skin and of the light sources.

Although well-illuminated test objects can rise above skin temperature, a rather high level of illumination is required. For example, a 100-watt incandescent source must be brought within 1 foot of a surface to irradiate it at a rate comparable with that produced by the juxtaposed skin at body temperature. If the temperature of the test objects did rise above that of the skin, the surface that had at first cooled the skin more rapidly and that had felt cooler (the more emissive surface) would afterwards heat the skin more rapidly and would feel warmer. Thus, a series of correct judgments could appear contradictory and inconsistent if the temperature of the test objects fluctuated above and below that of the skin, crossing threshold on each side. Any differences between the temperatures of the test objects induced by illumination would tend to bias the response in favor of warmer judgments for the emissive side.

A number of recent experiments on cutaneous discrimination between differently colored objects (Barrett & Rice-Evans, 1964; Buckhout, 1965; French, 1965; Youtz, 1966) have treated illumination as an important variable. The fact that the subjects were permitted to touch the test objects directly (except for the experiment of Barrett & Rice-Evans) suggests that the mechanism under investigation was the capacity of the test objects to store radiant heat for later transference to the skin by conduction. The interpretation of these experiments in most cases was complicated by: the effects of the shadow of the hand, the reflection of light to the skin by the test objects, heat conduction through an unknown area and duration of contact between the skin and the test objects, temporal changes in these variables accompanying hand movements, and the omission of information on irradiance and on the thermal properties of the test objects.

NOMENCLATURE

A number of names have been used to describe skin-mediated discriminations of the type normally requiring vision. Some reference to skin is appropriate to denote the sensitive organ and to distinguish between cutaneous and retinal functions. Reference to any particular skin layer, however, such as the dermis, is unduly specific. To differentiate between sensitivity of the type treated here from that which might imply a greater degree of spatial resolution, such terms as vision and sight should be eschewed for a term like sensitivity. Nor is the discrimination herein discussed analogous to color vision, for the multidimensional color space is compressed here into a single continuum, rate of heat exchange. But rate of heat exchange between observer and object is correlated with the hue of the object as well as with its lightness, so that a general term that subsumes both properties, such as color sensitivity, does serve to relate the sensory function to the visible differences between the objects discriminated. Because the discrimination actually depends upon thermal exchanges that are only statistically correlated with visible properties, however, perhaps cutaneous emissivity sensitivity (or ϵ sensitivity) would be a more accurately descriptive term.

IMPLICATIONS

Interest in cutaneous color sensitivity probably derives mainly from two sources: practical interest in exploiting new sensory abilities, particularly as they might benefit the blind; and theoretical interest in understanding the senses and the ways in which they respond to physical conditions.

It is difficult to imagine any direct, practical application of this sensory capacity. Even under optimum conditions it is a nearly threshold effect. There is no evidence even of the crudest spatial resolution. Holding the hand perpendicular to the boundary between the reflective and emissive halves of the test plate provides no cues as to which half is which; these cues only arise after the hand is rotated 90° and placed successively over the two halves of the plate. No convincing evidence of the more visual kind of cutaneous sensitivity is known to the writer.

Thus, the main interest properly centers on theoretical implications. The phenomenon reported by Youtz would have been most interesting if, after having been verified, it had pointed to new sensory properties or to flaws in the understanding of previously known properties. The demonstration, and its ready replicability, will establish the validity of a phenomenon at least very much like that reported by Youtz; the computations and temperature measurements, however, provide a reasonable explanation of the phenomenon that requires no new principles or sensory capacities. If this explanation accounts for all valid findings, then the phenomenon is of less interest than originally might have been expected.

Until new explanations are advanced, formal experiments in this area will be most easily interpreted if they provide the information necessary to determine the applicability of the present explanation to the results. A minimum in this respect is information on the temperatures of the skin and of all test objects, and on the emissivities of the test objects. Thermal control of the test objects would be of added value. Of further value would be information on the spectral absorption curves of the test objects, the spectral emission curves of all radiant sources, the irradiance from all sources as measured on the test objects and on the skin, the thermal properties of the test objects, and perhaps even the results of calorimetric measurements.

REFERENCES

BARNES, R. B. Thermography of the human skin. *Science*, 1963, **140**, 870–877.

BARRETT, S. M., & RICE-EVANS, P. Sensitivity of hands to visible light. *Nature*, 1964, **203**, 993.

BUCKHOUT, R. The blind fingers. *Perceptual and Motor Skills*, 1965, **20**, 191–194.

DAVIS, T. P. The heating of skin by radiant energy. In C. M. Herzfeld (Ed.), *Temperature, its measurement and control in science and industry: Part 3. Biology and medicine.* New York: Reinhold, 1961. Pp. 149–169.

DAVIS, T. P. In vivo temperature measurements. *Acta ophthalmologica*, 1963, Suppl. 76, 41–50.

FRENCH, C. N. Tactile-vision: Thermal and texture cues in the discrimination of black and white. *Nature*, 1965, **208**, 1352.

GARDNER, M. Dermo-optical perception: A peek down the nose. *Science*, 1966, **151**, 654–657.

GOLDBERG, I. M. On whether tactile sensitivity can be improved by exercise. *Soviet Psychology and Psychiatry*, 1963, **2**, 19–23.

HARDY, J. D. The radiation of heat from the human body: II. A comparison of some methods of measurement. *Journal of Clinical Investigation*, 1934, **13**, 605–614. (a)

HARDY, J. D. The radiation of heat from the human body: III. The human skin as a black-body radiator. *Journal of Clinical Investigation*, 1934, **13**, 615–620. (b)

HARDY, J. D., HAMMEL, H. T., & MURGATROYD, D. Spectral transmittance and reflectance of excised human skin. *Journal of Applied Physiology*, 1956, **9**, 257–264.

HARDY, J. D., & MUSCHENHEIM, C. The radiation of heat from the human body: IV. The emission, reflection, and transmission of infra-red radiation by the human skin. *Journal of Clinical Investigation*, 1934, **13**, 817–831.

HARDY, J. D., & MUSCHENHEIM, C. Radiation of heat from the human body: V. The transmission of infra-red radiation through skin. *Journal of Clinical Investigation*, 1936, **15**, 1–9.

HARDY, J. D., & OPPEL, T. W. Studies in temperature sensation: III. The sensitivity of the body to heat and the spatial summation of the end organ responses. *Journal of Clinical Investigation*, 1937, **16**, 533–540.

HARDY, J. D., & STOLL, A. M. Radiometric methods for measurement of skin temperature. *Methods in Medical Research*, 1954, **6**, 85–99.

HENDLER, E., & HARDY, J. D. Infra-red and microwave effects on skin heating and temperature sensation. *Institute of Radio Engineers Transactions on Medical Electronics*, 1960, 143–152. (ME-7)

HENDLER, E., HARDY, J. D., & MURGATROYD, D. Skin heating and temperature sensation produced by infra-red and microwave irradiation. In C. M. Herzfeld (Ed.), *Temperature, its measurement and control in science and industry: Part 3. Biology and medicine.* New York: Reinhold, 1961. Pp. 211–230.

HENSEL, H. Temperaturempfindung und intracutane Wärmebewegung. *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere*, 1950, **252**, 165–215.

HODGMAN, C. D., WEAST, R. C., & SELBY, S. M. (Eds.) *Handbook of chemistry and physics.* (42nd ed.) Cleveland: Chemical Rubber, 1960.

LELE, P. P., WEDDELL, G., & WILLIAMS, C. M. The relationship between heat transfer, skin temperature and cutaneous sensibility. *Journal of Physiology*, 1954, **126**, 206–234.

MOLNER, G. W., & ROSENBAUM, J. C., Jr. Surface temperature measurement with thermocouples. In C. M. Herzfeld (Ed.), *Temperature, its measurement and control in science and industry: Part 3. Biology and medicine.* New York: Reinhold, 1961. Pp. 3–11.

NAFE, J. P., & KENSHALO, D. R. Somesthetic senses. *Annual Review of Psychology*, 1962, **13**, 201–224.

NYUBERG, N. D. "Finger Sight"—the phenomenon of Roza Kuleshova. *Proceedings of the Federation of American Societies for Experimental Biology*, 1964, **22**, T701–T705. (Translation Suppl.)

OPPEL, T. W., & HARDY, J. D. Studies in temperature sensation: I. A comparison of the sensation produced by infra-red and visible radiation. *Journal of Clinical Investigation*, 1937, **16**, 517–524. (a)

OPPEL, T. W., & HARDY, J. D. Studies in temperature sensation: II. The temperature changes responsible for the stimulation of the heat end organs. *Journal of Clinical Investigation*, 1937, 16, 525–531. (b)

PICK, H. L., JR. Perception in Soviet psychology. *Psychological Bulletin*, 1964, 62, 21–35.

RANSON, S. W., & CLARK, S. L. *The anatomy of the nervous system, its development and function*. Philadelphia: Saunders, 1953.

WEDDELL, G., & MILLER, S. Cutaneous sensibility. *Annual Review of Physiology*, 1962, 24, 199–222.

YOUTZ, R. P. Aphotic digital color sensitivity. Paper read at Psychonomic Society, Bryn Mawr, Pa., September 1963.

YOUTZ, R. P. The case for skin sensitivity to color; with a testable explanatory hypothesis. Paper read at Psychonomic Society, Niagara Falls, N. Y., October 1964.

YOUTZ, R. P. Letter to the editor. *Scientific American*, 1965, 212, 8–9.

YOUTZ, R. P. Tactile color sensing under three intensities of illumination. Paper read at Eastern Psychological Association, New York City, April 1966. (Abstract)

ZOTTERMAN, Y. Thermal sensations. In J. Field, H. W. Magoun, & V. E. Hall (Eds.), *Handbook of physiology: Neurophysiology*. Vol. 1. Washington, D. C.: American Physiological Society, 1959. Pp. 431–458.

(Early publication received
November 9, 1965)

Psychological Review
1966, Vol. 73, No. 4, 295–305

TENSIONS IN PSYCHOLOGY BETWEEN THE METHODS OF BEHAVIORISM AND PHENOMENOLOGY[1]

NATHAN BRODY
Graduate Faculty of Political and Social Science
New School for Social Research

AND

PAUL OPPENHEIM
Princeton, New Jersey

3 loci of tension between behaviorism and phenomenology are discussed. Logical analysis indicates that the methods of phenomenology as defined are not relevant for the construction and test of psychological theory, although they may be valuable in other respects. 2 rapprochements are discussed in which complementary roles are assigned to behaviorism and phenomenology. Also, the problem of terminological rapprochement is discussed.

> Only silence before the *Thou* . . . leaves the *Thou* free. . . . Every response binds up the *Thou* in the world of *It*. That is the melancholy of man, and his greatness. For that is how knowledge comes about. . . . He will, indeed, have to grasp as an object that which he has seen with the force of presence, he will have to compare it with objects, establish it in its order among classes of objects, describe and analyze it objectively. Only as *It* can it enter the structure of knowledge [Buber, 1958, pp. 39–40].

In this paper, methodological differences between behavioristic and phenomenological psychology are discussed. The plan of the paper is as follows: *First,* a brief definition of behaviorism and phenomenology will show how these terms are conceived here. *Second,* an analysis will be made of three loci of tension between behaviorism and phenomenology. *Third,* the question of rapprochement will be discussed.

BEHAVIORISM AND PHENOMENOLOGY DEFINED

Behaviorism

Behaviorism is conceived here in Bergmann's (1956) sense of methodological behaviorism. According to this position, the fundamental data of psychology are the publicly observable behaviors of organisms; these are conceived as being describable in an intersubjective observation language. For the behaviorist, private experiences per se are not admissable as data, but only their publicly observable symptoms, if any. Behaviorists differ as to whether it is necessary or at least advisable to assign to private experience any role at all in scientific psychology. Behaviorists who assign a role to private experiences would do so by referring to them by means of theoretical (as opposed to observational) terms (see Eriksen, 1962; Jessor, 1958).

If a behaviorist wanted to introduce into the vocabulary of psychology theoretical terms referring to a person's private experience, it would be necessary for him to (*a*) define the terms by multiple convergent operations, that is, by several distinct operations which jointly determine the intended empirical meaning (see Garner, Hake, & Eriksen, 1956), (*b*) develop the theoretical relations between these terms and others, and (*c*) demonstrate the value of such terms for the explanation of behavior. These three conditions are not completely satisfied at present.

---

[1] The authors gratefully acknowledge critical remarks by C. G. Hempel. Thanks are also due to the Foundation for Research in Philosophy of Science, Princeton, New Jersey, for financial support. The Foundation is also supporting ongoing work on the relation between science and religion; this work suggested some of the ideas contained in this paper.

Phenomenology

Phenomenology is conceived in this paper in a sense similar to MacLeod's (1964) interpretation (cf. Edie, 1964, for a more general discussion of phenomenology and its relation to Existential philosophy). We shall define pure phenomenological psychology (ppp)[2] by its subject matter and by its method of investigation. The subject matter of ppp is all of experience. It is the intent of the ppp to study that which is immediately given in experience. MacLeod (1964) states in this connection:

> . . . colors and sounds are data; so are impressions of distance and duration; so are feelings of attraction and repulsion; so are yearnings and fears, ecstasies and disillusionments; so are all the relations—ranging from the crude and obvious to the delicate and intangible—with which the world presents us [p. 51].

The ppp studies the data of experience by a method which involves the suspension of implicit and explicit assumptions. (This method is reminiscent of what is called by Husserl, in a more technical context, "phenomenological reduction" and the act of "bracketing"—1962, especially pp. 10, 12, 96–100). That is, the ppp seeks to investigate experience freed, as far as possible, from the encumbrance of any preconceived theoretical ideas about that which is being investigated. This is sometimes referred to by Gestalt psychologists as naïve phenomenology (see Koffka, 1935, p. 73; Köhler, 1947). There is an attempt to investigate the immediate data given to the mind and not a theoretical construction or interpretation of these data. In the words of Lewis:

> There are in our cognitive experience two elements; the immediate data such as those of sense which are presented or given to the mind, and a form, construction, or interpretation, which represents the activity of thought. Recognition of this fact is one of the oldest and most universal of philosophic insights [1956, p. 38].

For a similar distinction made by phenomenologists between "fully reflexive categorical thought" and "prereflexive awareness," see Edie (1964). The subject matter of ppp is the former element.

Carried to its logical extreme, the method of ppp demands the bracketing of all concepts since they arise out of the activity of thought. It is therefore appropriate to call the type of experience involved in the method of ppp *nonconceptualized experience,* that is, an experience of the immediately given, sometimes referred to as "brute facts" (James, 1910) or "raw feels" (Feigl, 1958), not mediated by any concepts.

By characterizing ppp in this way, it is not implied that it is ever possible to have a purely nonconceptualized experience. But the method of ppp implies that one can approximate such an experience.

THREE LOCI OF TENSION BETWEEN PPP AND BEHAVIORISM

Construction and Evaluation of Psychological Theories

The first locus of tension involves the *context of discovery* (Reichenbach, 1938) for psychology, that is, the considerations that govern the way in which psychological theories and hypotheses are discovered. We deal here with what is at least a stylistic difference between ppp and behaviorism. If we restrict ourselves to the context of discovery, there is no reason why a behaviorist should not have the option of letting his explanatory efforts be initially guided by a phenomenological analysis of experience. He could then continue his investigation by collecting behavioral data. On the other hand, for the ppp, the former starting point is considered indispensable. This stylistic difference can become the basis for the evaluation of theories and in this sense be extended into the *context of justification* (Reichenbach, 1938). Behavioristic theories which fail to represent some aspect of experience discovered by the method of ppp have been criticized by the latter as inadequate (see MacLeod, 1964).

---

[2] We shall use ppp to also stand for "pure phenomenological psychologist" and rely on the context to make clear which meaning is intended.

Logical analysis of nonconceptualized experience. In order to evaluate such criticism, we shall present a logical analysis of nonconceptualized experience. This analysis is influenced in part by the writings of philosophers about what we call nonconceptualized experience (see Ayer, 1956; Black, 1954; Carnap, 1928, 1963; Lewis, 1946, 1956; Northrop, 1948, 1962). Also relevant are discussions of mysticism which can be considered a type of nonconceptualized experience (see Maslow, 1962; Stace, 1960; Zaehner, 1957).

What type of knowledge (the term knowledge is here understood in its usual conceptual sense) can be gained by an experiencer who has nonconceptualized experiences and who makes no attempt to introduce any hypothetical or theoretical construction into his discussion of his experiences? It is our contention that a person who restricts himself in such a way cannot have any knowledge at all.

In order for an experiencer to be able to claim that his experience has led to some sort of knowledge, he has to be able to give a description of the content of his experience for himself or others. Strictly speaking, any statement about a nonconceptualized experience involves an interpretation; this carries a conceptual element into the description. Let us examine more closely the types of inferences which are involved in such statements. Consider the statement "I saw red," or perhaps, "Redness was experienced by me," as an immediately given. We normally infer from such a statement that the experience of redness was not merely a figment of the experiencer's imagination. In addition, the statement is usually considered to imply that other persons, or the same person on another occasion, would experience redness if placed in the same situation. It is reasonably clear that such inferences would be logically excluded if one were restricted to reporting only what is immediately given in nonconceptualized experience. But can such a reporter legitimately affirm that he experienced redness? In order to assign properties such as red to an experience, it would be necessary to accept the division of the color continuum into the particular set of categories used in our everyday language. This classification carries a conceptual element into the description of the nonconceptualized experience (see Brown, 1965).

In general, whoever attempts to describe a nonconceptualized experience can do so only by a process of conceptualization which distorts and impoverishes the experience. Therefore, an experiencer who attempts to describe his nonconceptualized experience without such loss of content must in fact be speechless. In a similar vein, Black (1954) asserts that statements about the immediately given are not to be considered as descriptions but only as exclamations, as it were. Of course, a person can make a statement about his nonconceptualized experiences, but such statements are logically incommensurable with the experiences per se and can at best only be *correlated* with them, because statements per se involve the use of class concepts, whereas the idea of nonconceptualized experiences logically excludes the applicabiilty of class concepts. Northrop (1962, especially pp. 154, 237, 238) calls the relation between such incommensurables "epistemic correlation" (see also Havens, 1961).

The discussion of nonconceptualized experiences will now be summarized and partially extended by the presentation of a list of characteristics, each of which is logically implied by the meaning of nonconceptualized experiences. The characteristics assigned to such experiences may not be independent of each other.

1. During a nonconceptualized experience, the experiencer is not capable of distinguishing himself conceptually as experiencer from that which he is experiencing and is in this sense in *union* with it.

2. During a nonconceptualized experience, the experiencer is not capable of assigning a locus to his experience in a public spatio-temporal coordinate system. Thus, the experience could only have a locus in a private frame of reference with respect to space and time. With respect to time, the experience is called a *specious present* (see James, 1890, Ch. 15).

3. Nonconceptualized experiences are *ineffable* in the sense that during and after the nonconceptualized experience the experiencer has not at his disposal words which could be used by him to describe for himself or others the content of his experience. A fortiori, they are *propertyless,* since properties involving class terminology cannot be assigned to them. This implies that differences, if any, contained in the experience cannot be described; it remains in this sense an undifferentiated unity.

4. Nonconceptualized experiences are *solipsistic.* Since they cannot be verbally communicated to others, they have meaning, if any, only for the experiencer himself.

Two conclusions about the method of ppp. First, the advocates of ppp have criticized the method of analytic introspection, which had been used by the Structuralists, on the ground that their method was not a true investigation of experience, since the search for elements of experience leads to the introduction of elements where, in fact, the immediately givens are wholes. Thus, the method of investigation partially dictates its outcome (see Köhler, 1947; MacLeod, 1964). A similar argument can be made with respect to ppp: Its method emphasizes personal as distinguished from public time (see May, 1958, pp. 65–71). This is understandable as a logical consequence of its method (see Characteristic 2 above). In general, it will sometimes be difficult to know whether ppp has chosen a certain theme only as a logical consequence of its method of investigation, or in its own right because of its importance as an essential aspect of human experience, or as a result of confounding both elements.

Thus, paradoxically, the attempt to bracket all concepts in order not to bias the results of an investigation of experience introduces certain epistemic consequences which tend to influence the results of an investigation and creates a bias.

The second conclusion suggested by this analysis of nonconceptualized experiences is that these experiences do not provide the basis for the construction of any conceptual system. This conclusion follows directly from the ineffability of these experiences. Therefore, nonconceptualized experiences cannot be used as a basis for the criticism of any conceptual system. It is in this sense that we understand MacLeod's (1964) statement that ppp is only propaedeutic to the development of psychological theory.

Differences in Types of Explanation

The second locus of tension to be discussed here concerns the possibility of explanation in behaviorism and ppp. Explanation in science is generally thought to involve the use of some theory (Hempel & Oppenheim, 1948). Thus, strictly speaking, such a process of explanation is not compatible with the method of ppp.

Two possible alternatives can be adopted by the ppp: First, he can refuse to construct any explanatory system and rely rather on some kind of empathic or intuitive understanding which cannot be verbally communicated at all. Some Existential psychoanalysts have apparently used this approach. For example, Tymieniecka (1962, p. 69) mentions three fundamental "empirical" statements advanced by Jaspers. These are:

1. Man in his entirety cannot be objectively (conceptually) known;
2. Cognition is not limited to causal explanation;
3. There are essential aspects of mind which are by their very nature not accessible to experimental methods and cannot be rationally grasped.

Second, the ppp can at one time take the approach mentioned above. At another time, he can transcend the phenomenal world and construct a system which is conceptual and which is believed to preserve some of the intuitive insights obtained in nonconceptualized experience. But in such a case, ppp has only been a heuristic device limited to the context of discovery. It is not clear how the insights obtained by these methods can be carried over into the actual construction of theories.

Methods of Test

The third locus of tension concerns the method of test. For the behaviorist, any psychological theory, law, or statement must be ultimately testable with reference to intersubjectively observable symptoms of phenomena. The advocates of ppp, in so far as they do not exclude any theoretical statements, may differ as to their agreement with this demand of the behaviorist. For example, Rogers (1964) would agree with this demand. But whoever accepts this demand becomes in effect a methodological behaviorist, since his conceptual system rests ultimately on a behavioristic base, and ppp is used only in the context of discovery. Other advocates of ppp would reject any such insistence.

What role, if any, can be assigned to nonconceptualized experiences in the testing process? At one time or another, several philosophers have suggested that nonconceptualized experiences provide the basis for testing all scientific statements (see Carnap, 1928, p. 273; Lewis, 1946, 1956; Russell, 1912, Ch. V; Schlick, 1934).

To our knowledge, no one has ever presented in sufficient detail a logical analysis of such a testing process. We shall discuss an approach by Schlick (1934) because he comes very close to this desideratum: He takes his point of departure from the idea that the test of a scientific statement always ultimately involves the derivation of some statement describing some observable phenomenon. Schlick, in common with other philosophers of science, calls the sentences in which the results of observations of this kind are reported "protocol sentences." Such a sentence may for example have the form: "John Doe perceived at that and that time, at that and that place, this and that." Protocol sentences may thus provide a basis for testing scientific statements, but not an absolutely certain one. They are hypothetical; it is conceivable that despite great care the sentence written down by an observer in the laboratory protocol is false.

Schlick, however, argues that there is another kind of sentence which provides a more suitable basis for the test of scientific sentences than protocol sentences. He calls them "Konstatierungen" and characterizes them as follows: They are descriptions of what is immediately experienced and always have the demonstrative form "here now this," for example, "here now blue," without any reference to an experiencer. Konstatierungen can be known with absolute certainty, though only in the fleeting moment of the occurrence of the experience mentioned in the Konstatierung. In addition, Konstatierungen and the certainty which attaches to them are private. A Konstatierung cannot be written down nor be communicated in words to others because, for lack of general definitions, the meaning of the words "here," "now," "this" can only be indicated by demonstrative gestures, like pointing, and not in words.

Thus, Schlick's Konstatierung can be understood as a description of a nonconceptualized experience by an experiencer during the moment of his experience. We shall revert presently to the difficulties of assigning any role to Konstatierungen in the testing process of scientific sentences. (For critical comments about Schlick's paper, see Ayer 1953, especially pp. 10, 11, 90, 91; Nagel, 1954, p. 219.)

Despite the importance Schlick assigns to Konstatierungen in the testing process, they cannot, for him, be the basis for the construction of theories, because Konstatierungen are only momentary and have passed when one starts the construction.

A position similar to Schlick's, as to the role of nonconceptualized experiences in the testing process, has been taken by Lewis. He states that "without . . . apprehensions of direct and indubitable content of experience, there could be no basis for any empirical judgment, and no verification of one [1946, p. 182]." In his book, *Mind and the World Order,* Lewis (1956, Chs. 2–5) discusses this in detail. He calls the qualitative characteristics of an immediately given "qualia." They have no names as distinguished from the properties of objects. For Lewis, absorption in the immediately given transcends conceptual thinking altogether, whereas knowledge of objects involves conceptual interpretation. Though the experience of a quale is ineffable, Lewis (1956, p. 125) takes as an example the statement, "this is the same ineffable 'yellow' that I saw yesterday." Speaking in this way about the "ineffable yellow" seems like a contradiction in terms, and, interpreting it as the same as yesterday, contradicts, in our opinion, characteristics which Lewis assigns, as we do, to nonconceptualized experiences, namely, that they are not only ineffable, but propertyless and private as well.

Unless these contradictions are resolved, we do not see how Lewis is entitled to state that "without the correlation of concept and qualia no experience could verify or fail to verify anything [p. 144]."

The treatment of the testing process presented by both Schlick and Lewis is not compatible with the logical analysis presented in this paper: First, the question of the certainty of nonconceptualized experiences is a controversial one in philosophy (see Allport, 1960, p. 139; Ayer, 1956; Black, 1954; Born, 1924, pp. 4–5; Goodman, 1952; Tymieniecka, 1962, pp. 3–16; Weyl, 1949, p. 116). In addition, whatever position one takes about certainty, it is clear that in order to test a scientific statement by a nonconceptualized experience, or by a Konstatierung which purports to be a description of such an experience, the experiencer would have to abstract with reference to some relevant property of the experience. Otherwise, he could not know whether or not the content of his experience confirmed or disconfirmed the statement. Let us consider an example: Suppose an experiencer were to test a sentence referring to the occurrence of a coincidence, for example, between the pointer of an instrument and a mark on a scale. The experiencer might then have an experience which he might reasonably conceptualize, at a later time, as a coincidence. However, during the moment of his nonconceptualized experience, he cannot know whether what he is experiencing is or is not what, in the sentence, is meant by a coincidence, unless he makes a conceptual interpretation of his experience. If the experiencer, during his nonconceptualized experience, *could* know whether what he is experiencing is a coincidence, then he should be able to communicate his experience to others. Conversely, that which cannot be described to others cannot be described to oneself in terms which are clearly understandable to others. We do not see how Lewis or Schlick could overcome this objection. Thus, on the view presented here, nonconceptualized experiences per se cannot be used to test scientific statements.

As mentioned previously, ppp cannot be used as a basis for the construction of theories, nor, as we see now, for the purpose of testing any scientific statement, and ppp cannot even be used as a means of describing experience. Thus, on the view presented here, ppp is not relevant per se for scientific psychology. However, the analysis may have some relevance for understanding other important aspects of ppp.

As mentioned above, any conceptualization of nonconceptualized experience can lead to a feeling that the conceptual statement has impoverished and distorted the experience. The feeling can be particularly strong, for example, with respect to aesthetic experiences, where we may often feel the sheer inability of language to capture the subtlety of such an experience (see Schachtel, 1951). This feeling of impoverishment and distortion may be accompanied by a feeling of disunion between, or separation from, experiencer and that which is experienced. These feelings can become particularly acute when the object of our concern is another human being. Conceptualization, by assigning properties to a person, is sometimes even considered as a dehumanization. Such concerns are prevalent among Existential psychoanalysts and explain in part their rejection of Freud (see May, Angel, & Ellenberger, 1958).

We have quoted Buber at the beginning of this paper because it seems to us that he has clearly understood the advantages and disadvantages inherent in the act of conceptualization. Expressed in Buber's (1958) terminology, an experiencer has the following choice between two modes of relation: Either he may choose to have, in a receptive moment, a direct experience of another human being, a *Thou*, without separation by intervening concepts. But there can be no knowledge in the *I-Thou* relation. Or, he may choose to have knowledge, to describe, and analyze objectively. But in this *I-It* relation, man experiences another human being in separation from him.

A position similar to Buber's has been taken by Tillich (1960). He argues, in effect, that certain nonconceptualized experiences though not valuable for science at all are valuable in another way, namely, as a basis for religion. Similarly, on allegedly empirical grounds, Maslow (1962) has argued that peak experiences (a form of nonconceptualized experience) are valuable in that they promote psychological growth and health. What is maintained here is that any advantages inherent in nonconceptualized experiences are obtained at the price of their irrelevance per se for science. Also, the view presented here should not be interpreted as implying that one should not or cannot take nonconceptualized experiences as an object of scientific investigation. For example, the study of mysticism represents an interesting and legitimate scientific endeavor.

QUESTION OF RAPPROCHEMENT

Despite the fact that the analysis presented here does not permit any role assignment to ppp in scientific psychology, it is of some interest to examine the difficulties confronted by various approaches to rapprochement which would be faced by those psychologists who would not entirely accept the logical analysis presented here.

Complementary Roles for ppp and Behaviorism

A possible basis for rapprochement between ppp and behaviorism can be developed by assigning to them complementary roles in a unified psychology. We will examine two such possible bases for rapprochement.

A first basis would be afforded by the following two principles:

1(*a*). Only the methods of the behaviorist (and not those of the ppp) can be used to construct theoretical systems.

1(*b*). Only the methods of the ppp (and not those of the behaviorist) can be used to test theories. (Of course, 1*b* is contrary to the position taken in this paper.)

To our knowledge, such a basis for rapprochement has never been suggested by any psychologist. It is compatible with the position of such philosophers as Schlick and Lewis. It is not at all clear that the ppp would be satisfied with this abstention from theorizing as suggested by 1*a* above. As for condition 1*b*, most behaviorists would probably find this rather controversial. The behaviorists' insistence on an intersubjectively observable base for testing statements and their demand for observations which result in a high degree of intersubjective agreement tend to restrict the class of statements which could be directly tested to protocol sentences about publicly observable symptoms. No such restriction is implied by a method of test based on nonconceptualized experiences. In general, the class of statements which can be directly tested by nonconceptualized experiences appears to be larger than the class of statements which refers to experiences for which high degrees of intersubjective agreement can be obtained. Thus, if one can allegedly experience emotions as immediately given, then one could test directly statements about emotions. If one believes the mystics and accepts condition 1*b*, it would be possible to have a nonconceptualized experience of the divine, thus providing a basis for the direct test of statements about the divine. If nonconceptualized experiences are used as a means of testing statements instead of intersubjective tests by behavioristic symptoms, not only is the rational reconstruction of the testing process changed, but also the class of statements directly testable becomes larger. Of course, it is possible to add to condition 1*b* an arbitrary restriction which would vitiate the behaviorist's probable objection to the enlargement of the class of testable statements. Namely, it is possible to restrict the use of nonconceptualized experiences to the ultimate test of those statements which can be directly tested by behavioristic procedures, for example, to protocol sentences. If this is done, the class of statements which are directly testable is not enlarged. However, the addition of such a restriction would violate the intent of the ppp to study all of experience and to afford equal significance to all aspects of experience. Such a restriction would, in effect, divide the class of nonconceptualized experiences into two subclasses, one of which is assigned the role of testing scientific statements, and the other is not assigned any such role at all.

A second basis for rapprochement follows below:

2(*a*). Only the methods of ppp (and not those of the behaviorist) can be used to construct theoretical systems. (Of course, 2*a* is contrary to the position taken in this paper.)

2(*b*). Only the methods of the behaviorist (and not those of the ppp) can be used to test theoretical systems.

This is the basis of rapprochement suggested by Rogers (1964). The principal difficulty in connection with this basis for rapprochement involves the behaviorist's possible objection to condition 2*a*. Such a condition apparently restricts the theorist's ability to freely postulate theoretical concepts. Einstein (as cited by Holton, 1965 [3]) has expressed this idea as follows:

> Everything conceptualizable is constructive and not derivable in a logical manner from immediate experience. Therefore we are in principle completely free in the choice of those fundamental conceptions upon which we found our rendition of the world. Everything depends only on this: to what extent our construction is suitable for bringing order into the apparent chaos of the world of experience.

Condition 2*a* implies that the theorist is permitted to introduce only those concepts which somehow have a foundation in what is given.

Terminological Rapprochement

Any rapprochement between ppp and behaviorism would be dependent on agreement on terminology. There are surely terms in the vocabularies of ppp and behaviorism which could be successfully coordinated. Terms referring to certain sensory experiences, for example, "red," can be coordinated and assigned compatible meanings for both ppp and behaviorism. Other terms, the theoretical status of which has not been extensively developed and which are close to ordinary experiences, might also be assigned compatible meanings, for example, terms referring to emotional states, such as reverence. However, the fulfillment of this condition is dependent on the progress of psychological theory. As behavioristic psychology develops, terms which are used in a relatively atheoretical way, for example, "fear," acquire modified meanings by virtue of having entered into a network of theoretical laws and relations which serve to specify the meaning of the term. It is difficult to imagine how the modified meaning of such terms can be directly experienced or coordinated in any sense with an immediately given experience. A direct coordination and compatibility of meanings is possible only after their reduction to directly observable symptoms. Another obstacle to the fulfillment of a terminological rapprochement consists in the introduction of terms into a behavioristic psychology which have no counterpart at all in nonconceptualized experience, for example, terms contained in a highly formalistic mathematical model. Thus, as behavioristic psychology develops, the possibility of rapprochement between ppp and behaviorism may grow more remote in this respect. On the other hand, it can happen that theoretical psychology will develop in such a way that there will remain many aspects of nonconceptualized experience which have little counterpart in the terminology of behavioristic psychology, for example, the experience of disgust is at present without any extensively developed counterpart in the theoretical language of behavioristic psychology (see MacLeod, 1964).

---

[3] This quote is an English translation by Gerald Holton contained in the paper presented by him at the 11th International Congress of the History of Science, Warsaw and Krakow, August 1965. The German text is contained in Einstein's unpublished papers.

There is one way in which future developments within behavioristic psychology might lead to a greater degree of terminological agreement between ppp and behaviorism. As Zener (1958) has pointed out, the insistence that science deals only with statements about intersubjectively valid observations must be qualified if it is to be properly understood. This statement should not be taken to imply that only statements about observations where the intersubjective agreement among a large class of observers is almost perfect should be considered as data in science. Since the degree of agreement with respect to protocol sentences is never perfect, each psychologist is faced with a somewhat arbitrary decision about the degree of agreement with regard to some observation that he will demand. Zener would, for example, have the psychologist study reports about aesthetic qualities of experience which might be based on a less than perfect agreement among a small group of individuals with refined aesthetic sensibilities. To the extent to which behavioristic psychology is open to reports about subtle qualities of experience and is willing to accept them as data to be explained, there should be an increased possibility of terminological agreement between ppp and behaviorism.

REFERENCES

ALLPORT, G. *The individual and his religion.* New York: Macmillan, 1960.

AYER, A. J. *Language, truth and logic.* New York: Dover, 1953.

AYER, A. J. *The problem of knowledge.* Harmondsworth, England: Penguin, 1956.

BERGMANN, G. The contribution of John B. Watson. *Psychological Review,* 1956, **63,** 265–276.

BLACK, M. The language of sense-data. In M. Black (Ed.), *Problems of analysis.* Ithaca: Cornell University Press, 1954. Pp. 58–79.

BORN, M. *Einstein's theory of relativity.* London: Methuen, 1924.

BROWN, R. *Social psychology.* New York: Free Press of Glencoe, 1965.

BUBER, M. *I and thou.* New York: Scribners, 1958.

CARNAP, R. *Der logische Aufbau der Welt.* Berlin, Germany: Im Weltkreis-Verlag, 1928. (Republished: *Logical reconstruction of the world and problem of appearance in philosophy.* Berkeley: University of California Press, 1966.)

CARNAP, R. Nelson Goodman on "Der logische Aufbau der Welt." In P. A. Schlipp (Ed.), *The philosophy of Rudolf Carnap.* LaSalle, Ill.: Open Court, 1963. Pp. 944–947.

EDIE, J. M., III. Recent work in phenomenology. *American Philosophical Quarterly,* 1964, **1,** 115–128.

ERIKSEN, C. W. (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962.

FEIGL, H. The "mental" and the "physical." In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science.* Vol. 2. Minneapolis: University of Minnesota Press, 1958. Pp. 370–497.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychological Review,* 1956, **63,** 149–159.

GOODMAN, N. Sense and certainty. *Philosophical Review,* 1952, **61,** 160–167.

HAVENS, J. The participant's vs. the observer's frame of reference in the psychological study of religion. *Journal for the Scientific Study of Religion,* 1961, **1,** 79–87.

HEMPEL, C. G., & OPPENHEIM, P. Studies in the logic of explanation. *Philosophy of Science,* 1948, **15,** 135–175.

HUSSERL, E. *Ideas.* New York: Collier Books, 1962.

JAMES, W. *The principles of psychology.* Vol. 1. New York: Holt, 1890.

JAMES, W. *The meaning of truth.* New York: Longmans, Green, 1910.

JESSOR, R. The problem of reductionism in psychology. *Psychological Review,* 1958, **65,** 170–178.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

KÖHLER, W. *Gestalt psychology.* New York: Liveright, 1947.

LEWIS, C. I. *Knowledge and valuation.* LaSalle, Ill.: Open Court, 1946.

LEWIS, C. I. *Mind and the world order.* New York: Dover, 1956.

MACLEOD, R. B. Phenomenology: A challenge to experimental psychology. In T. W. Wann (Ed.), *Behaviorism and phenomenology.* Chicago: University of Chicago Press, 1964. Pp. 47–78.

MASLOW, A. H. *Toward a psychology of being.* Princeton: Van Nostrand, 1962.

MAY, R. *Contributions of existential psychotherapy.* In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence.* New York: Basic Books, 1958. Pp. 37–91.

MAY, R., ANGEL, E., & ELLENBERGER, E. F. (Eds.) *Existence.* New York: Basic Books, 1958.

NAGEL, E. *Sovereign reason.* Glencoe, Ill.: Free Press, 1954.

NORTHROP, F. S. C. *The logic of the sciences and humanities.* New York: Macmillan, 1948.

NORTHROP, F. S. C. *Man, nature and God.* New York: Simon & Schuster, 1962.

REICHENBACH, H. *Experience and prediction.* Chicago: University of Chicago Press, 1938.

ROGERS, C. R. *Toward a science of the person.* In T. W. Wann (Ed.), *Behaviorism and phenomenology.* Chicago: University of Chicago Press, 1964. Pp. 109–140.

RUSSELL, B. *The problem of philosophy.* London: Oxford University Press, 1912.

SCHACHTEL, E. On memory. In P. Mullahy (Ed.), *A study of interpersonal relations; new contributions to psychiatry.* New York: Grove Press, 1951. Pp. 1–49.

SCHLICK, M. Ueber das Fundament der Erkenntnis. *Erkenntnis,* 1934, **4,** 79–99.

STACE, W. T. *Mysticism and philosophy.* Philadelphia: Lippincott, 1960.

TILLICH, P. The relationship today between science and religion. In J. A. Clark (Ed.), *The student seeks an answer.* Waterville: Colby College Press, 1960. Pp. 297–306.

TYMIENIECKA, A. T. *Phenomenology and science in contemporary European thought.* New York: Noonday Press, 1962.

WEYL, H. *Philosophy of mathematics and natural science.* Princeton: Princeton University Press, 1949.

ZAEHNER, R. C. *Mysticism sacred and profane.* London: Oxford University Press, 1957.

ZENER, K. The significance of experience of the individual for the science of psychology. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science.* Vol. 2. Minneapolis: University of Minnesota Press, 1958. Pp. 354–369.

(Early publication received January 26, 1966)

Psychological Review
1966, Vol. 73, No. 4, 306–326

DESCRIPTIVE BEHAVIORISM VERSUS COGNITIVE THEORY IN VERBAL OPERANT CONDITIONING[1]

CHARLES D. SPIELBERGER AND L. DOUGLAS DENIKE
Vanderbilt University *University of Southern California*

The relative adequacy with which descriptive behaviorism and cognitive learning theory can account for the findings of verbal conditioning experiments was examined along with the methodological and epistemological assumptions on which these interpretations are based. Methodological problems associated with the techniques employed for the assessment of mediating cognitive states (awareness) were discussed, and evidence was presented to support the view that "what is learned" in verbal conditioning is awareness of a correct response-reinforcement contingency. The implications of verbal conditioning findings for the more general question of learning without awareness were considered, and it was concluded that the investigation of awareness in relation to other behavior is requisite to the development of an adequate theory of human learning.

Operant behavior generally refers to responses which an organism in a given environment displays spontaneously without special training. Operant conditioning usually implies a variety of experimental techniques wherein a subject is rewarded after engaging in operant behavior of a selected kind. The successful conditioning of an operant response is inferred from an increase in its rate of occurrence as a function of reward (reinforcement) administered by the experimenter. Through the efforts of those who have used operant conditioning procedures, significant advances in our knowledge of animal learning have been made possible, and widespread interest has been generated in the use of such procedures for investigating human learning. The purpose of the present paper is to examine theoretical and methodological issues which have arisen in recent applications of operant conditioning techniques to verbal learning.

In verbal operant conditioning, hereafter termed verbal conditioning, the subject is typically instructed to speak in accordance with a particular task. He is not told that this task involves reinforcement or learning. The experimenter attempts to change the rate of emission of certain responses by the systematic application of a reinforcing stimulus, usually some form of social reward, such as the experimenter saying "Mmm-hmm" or "Good." Studies of verbal conditioning originated in close connection with the concepts and procedures developed by Skinner (1938) with infrahuman species. Greenspoon, a pioneer in verbal conditioning research, maintains that

> . . . it should be possible to work with verbal behavior in much the same way that experimenters have worked with the behavior of rats, pigeons, etc. It should also be possible to investigate the same kinds of variables that have been investigated with the non-verbal behavior of humans and infrahumans [1962, p. 511].

---

[1] Work on this paper was supported in part by grants to the first author from the National Institutes of Mental Health (MH 7446) and Child Health and Human Development (HD 947), United States Public Health Service. We are indebted to Norman Cliff, Jum C. Nunnally, Henry Slucki, and Donald L. Thistlethwaite for their critical comments on the manuscript. The first author is now at the National Institute of Mental Health, Bethesda, Maryland.

Thus, the Skinnerian (1957, 1963a, 1963b) approach has been clearly apparent both in the methods employed and the variables investigated in most verbal conditioning experiments.

In keeping with Skinner's emphasis on the "experimental analysis of behavior," many verbal conditioners have largely ignored the possibility that subjects' awareness of response-reinforcement contingencies might influence their conditioning performance. Investigators who have examined subjects' verbal reports in relation to their conditioning performance have discovered considerable evidence suggesting that cognitive processes mediate performance gains in verbal conditioning. But descriptive behaviorists have been quick to rejoin that the relationships between awareness and performance observed by cognitively oriented researchers have arisen from artifacts associated with the verbal report procedures from which awareness was inferred. Thus, in verbal conditioning, broadly opposed theoretical systems lock horns: Descriptive behaviorists argue for learning without awareness, those of cognitive persuasion argue against it, and proponents of each point of view generate methodological criticisms of the experimental work carried on in the opposing camp.[2]

It would appear that radically different epistemologies underlie the theoretical differences between Skinnerian and cognitive researchers. The epistemological issues, which revolve around the admissibility of conscious awareness as a desideratum for psychological science, have been discussed in detail elsewhere (Spielberger, 1965). Therefore, rather than restating these issues here, a simple analogy may better serve to illustrate how varying pretheoretical assumptions about the general nature and scope of science can lead to highly disparate approaches to a concrete scientific problem.

It seems reasonable to conjecture that biologists of an earlier generation who did not believe in protozoa probably contended at times that objects too small to be observed by the naked eye were *irrelevant to scientific analysis* (of disease, for instance). Attempts to explore the domain of the microscopic world might have been disdained by such observers as revealing only illusory effects traceable to light diffusion in the microscope, and so on. Similarly, we might expect that present-day psychologists who consider thoughts and hypotheses to be beyond the limits of "scientific" inquiry would not vigorously search for them in experimental subjects. Furthermore, we should not be surprised to find that such psychologists were uninterested and unskilled in evaluating cognitive phenomena which for them do not exist. On the other hand, early biologists who believed in protozoa might occasionally have "seen" them when they were not there. Similarly, we might expect, as has been suggested by Farber (1963), that cognitively oriented verbal conditioning researchers, in their eagerness to attribute behavior to mediating conscious processes, would sometimes inadvertently suggest (or erroneously infer) awareness in questioning subjects who show performance gains.

---

[2] It should be noted that S–R behavior theorists in the Hull-Spence tradition are not a party to this controversy. The model developed by these theorists was designed to account for the behavior phenomena exhibited by nonarticulate organisms or by humans in simple learning situations in which the operation of higher mental processes was minimal, for example, in eyelid conditioning and rote learning. A major difference between the views of descriptive behaviorists and those of Hull and Spence is that the latter never claimed that their theoretical concepts would hold for complex verbal processes. In his discussion of the postulates and methods of behaviorism nearly 2 decades ago, Spence (1948, p. 76) noted that:

> . . . in dealing with the more complex types of animal and human behavior, implicit emotional responses, covert verbal responses and not easily observable receptor-exposure and postural adjustments will have to be postulated . . .

and that:

> It is in this realm of theorizing that the verbal reports of human subjects are likely to be of most use to the behavior theorist, for presumably these reports can be made the basis on which to postulate the occurrence of these inferred activities.

One obvious implication of the above analogy is that competing scientific theories lead to different experimental procedures and often to observational errors which support the theoretical predilections of the investigator (Rosenthal, 1963). In most cases, however, such methodological differences and observational errors are eventually resolved since competing theories generally lead to more sensitive experiments. Furthermore, good experiments contribute to the accumulation of a composite set of facts which facilitate the convergence of theoretical schools and the establishment of an organized body of scientific knowledge (Campbell, 1963). But, if the epistemological assumptions which underlie competing scientific theories differ, the methodological consequences of such differences may lead to the collection of noncomparable data about which pointless theoretical controversy is generated. Unfortunately, this seems to be the case in verbal conditioning.

In this paper, we propose to examine the relative merits of cognitive and Skinnerian interpretations of verbal conditioning and the methodological assumptions on which these interpretations are based. The organization of the paper is as follows: In Section I the use of verbal report measures as indexes of awareness (mediating cognitive processes) will be discussed, and the various objections which investigators of a Skinnerian bent have raised concerning these measures will be enumerated and analyzed. In Section II methodological problems associated with the assessment of awareness from postconditioning interviews will be considered, and procedures for assessing awareness during conditioning will be described along with the results of two experiments in which such procedures were utilized. A theoretical analysis of verbal conditioning in terms of cognitive concepts will be presented in Section III. In Section IV the significance of the verbal conditioning paradigm for the more general question of learning without awareness will be taken up.

I. CAN VERBAL REPORTS BE INTERPRETED AS INDEXES OF MEDIATING COGNITIVE PROCESSES?

Methodological criticism has been associated with attempts to condition verbal behavior since Thorndike's early experiments in this area. In these Thorndike was interested in determining whether the law of effect operated when the subject was unaware of the contingency of reinforcement. Utilizing a word-association task, Thorndike and Rock (1934) reinforced sequential as opposed to denotative associations (OVER-the hill, as opposed to OVER-above). They reasoned that if subjects gained sudden insight (awareness) into the response-reinforcement contingency, the number of sequential responses which they gave would rise sharply. Hence, they regarded the gradual increase in sequential responses, which they in fact obtained, as providing evidence for learning without awareness. This interpretation, however, was immediately challenged by the demonstration that similar gradual performance gains ensued even when subjects were informed of the basis of reinforcement (Irwin, Kauffman, Prior, & Weaver, 1934). It is perhaps an unfortunate commentary on the narrow, behavioristic zeitgeist then prevailing that the simple procedure of questioning subjects was not utilized in either of these studies. Irwin et al. (1934), while they did not fill the need for measures of insight or awareness that are operationally independent of performance, helped at least to make the need evident. More recently, the quest for such measures has generally led to the use of systematically obtained verbal reports.

Now Skinnerian verbal conditioners do not hold that subjects should never be asked questions, nor do they assert that reliable relationships between verbal reports and performance measures cannot be found. They do argue that such relationships when found do not constitute adequate evidence that awareness plays a *causal* intervening role. However, descriptive behaviorists have not attempted to investigate the adequacy of verbal report procedures. Thus, their criticisms of them appear to stem chiefly from the implicit assumption that consciousness is in principle unknowable and/or that it is unnecessary to consider it. These epistemological biases are not often made explicit, but under their influence criticisms of cognitive interpretations of verbal conditioning phenomena are made on methodological and theoretical grounds. It is to such criticisms[3] that we now turn.

1. The first group of objections advanced by descriptive behaviorists to the interpretation of verbal reports in cognitive terms is concerned with the possibility that awareness is *suggested* by the procedures which are employed to assess it. Questioning techniques used to evaluate awareness may, it is argued, differentially suggest the contingency of reinforcement to subjects who condition as compared to those who do not. When both the conditioning task and the awareness interview are conducted by the same experimenter, it is certainly plausible that the experimenter might inadvertently ask questions and/or record answers in such a way as to suggest and/or impute awareness to those subjects who showed performance gains. A related way in which the interviewer might artifactually bring about a relationship between performance and awareness is through cues provided by the wording of interview questions. Questioning subjects about their experiences during an experiment obviously leads them to reflect on their performance and possibly also to develop hypotheses about the response-reinforcement contingency.

2. A second general class of criticism contends that subjects condition without awareness, notice their increasing outputs of the critical response class, and then *label* the response-reinforcement contingency. In other words, it is argued that the subjects' knowledge of the contingency emerges as a result of an automatic conditioning process, and those subjects who indicate in a postconditioning interview that their performance was mediated by their hypotheses are merely rational-

---

[3] These criticisms have not been listed and treated in detail in any single article. What follows is an attempted coverage of relevant objections to interpretations of verbal reports in verbal conditioning in terms of mediating states. These have been adapted from friendly and unfriendly sources, including Dulany (1961, 1962), Eriksen (1958, 1960, 1962), Greenspoon (1962, 1963), Kanfer and his colleagues (Kanfer & Marston, 1961, 1962; Kanfer & McBrearty, 1961), Krasner (1962), Krasner and Ullmann (1963), Postman and Sassenrath (1961), Salzinger (1959), Spielberger (1962, 1965), and especially Dulany (1963).

izing their performance gains. This line of criticism calls into question correspondences between performance and awareness based on measures taken after performance gains have ensued and points up the theoretical significance of the temporal sequence of events in verbal conditioning, that is, whether in fact performance gains precede awareness or occur subsequent to the subjects' development of awareness.

3. A more complicated variant of the point of view indicated in 2 above has been suggested by Postman and Sassenrath (1961)[4] who state that "awareness as reflected in verbalization may represent an advanced stage in the development of a habit under conditions of reinforcement [p. 124]." They go on to suggest that the verbalization of a response-reinforcement contingency "*may be considered at the same time a result of past improvement and a condition for futher improvement* [p. 124]." The tenability of this formulation, which ascribes a secondary mediational role to awareness, rests upon the demonstration of significant amounts of learning prior to the subjects' verbalization of a correct or partially correct contingency.

4. Another line of criticism contends that awareness is merely a covert response which is conditioned over trials, in an automatic fashion, at the same time as the reinforced response class is strengthened. Postman and Sassenrath (1961), in discussing symbolic mediators, have posited this sort of mechanism to explain the acquisition of hypotheses in instrumental learning. They state: "The differential reinforcements administered for the overt instrumental responses will at the same time selectively strengthen the correct mediating responses [p. 132]." In this view, awareness is assigned the properties of an operant response, and it is hypothesized that covert verbalizations of the reinforcement principle (awareness) are directly and automatically strengthened by the same reinforcement process through which other instrumental responses are conditioned.[5]

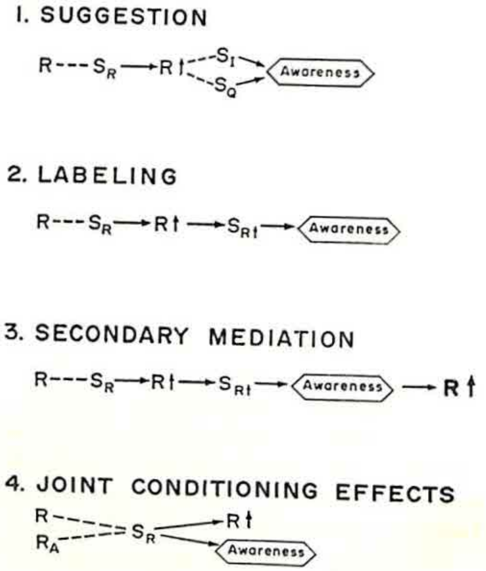

FIG. 1. Schematic diagrams depicting noncognitive interpretations of verbal reports of awareness in verbal conditioning.

The formulations described above are presented as schematic diagrams in Figure 1. These diagrams are not intended to represent particular viewpoints in detail, but are offered only tentatively as an aid in expressing interpretations of verbal conditioning which consider verbal reports primarily as responses rather than as indexes of mediating states. In each schematic, the temporal sequence of events is from left to right. Events which are merely temporally contiguous are indicated by dotted lines; presumed causal relationships between events are represented by solid arrows.

---

[4] It should be noted that although Postman and Sassenrath cite findings of verbal conditioning investigations as relevant to this interpretation, their conclusions are based primarily upon Thorndikian experiments in which gradual increments in performance typically precede the point of verbalization.

[5] A similar interpretation is suggested by Kanfer and McBrearty (1961) who contend that:

> . . . verbalization of the response-reinforcement contingency may be independently affected by *different* antecedent conditions than the acquisition rate of the (reinforced verbal) response class, even though experimental conditions may sometimes facilitate an interaction between these two events [p. 116].

Schematic 1 illustrates the two possibilities whereby awareness might be suggested by questioning procedures to subjects who showed performance gains. The symbol R stands for the subject's verbalization of responses belonging to the reinforced response class. S_R refers to the reinforcing stimulus administered by the experimenter for each R. $R\uparrow$ indicates performance gains.. S_I represents biasing stimuli that might be transmitted unwittingly to the subject by an interviewer who had knowledge of the subject's conditioning performance. S_Q stands for possible cues conveyed by the wording of questions which might lead subjects who showed performance gains to construct correct hypotheses after the fact by reflecting on their performance.

Schematic 2 diagrams the hypothetical automatic strengthening-labeling sequence in which subjects are said to condition without awareness and then to become aware of the response-reinforcement contingency after noticing their increasing output of reinforced responses, $S_{R\uparrow}$.

Schematic 3 depicts the hypothetical operation of mediating processes subsequent to automatic conditioning. As described in Schematic 2, the subject becomes aware as a result of increments in performance brought about by the automatic effects of reinforcement and utilizes this awareness further to enhance performance gains. In this diagram, $R\uparrow$ symbolizes augmented performance consequent upon the subject becoming aware of the correct contingency.

Schematic 4 illustrates the view that reinforcement acts simultaneously to condition both the reinforced response class and verbalization of the reinforcement principle. Presumably, reinforcement would operate in this paradigm to strengthen covert verbalizations of the correct response-reinforcement contingency, R_A.

Underlying all of the above formulations is the epistemological viewpoint mentioned earlier. It would appear that what descriptive behaviorists object to basically is the use of verbal reports as a Trojan horse for the reintroduction of private experiences into psychology. Since private experience does not meet positivistic canons of interobserver reliability, they argue largely on epistemological grounds that interest in it remains but a vestige of discredited introspectionism. However, the heuristic value of operationally defined awareness variables in verbal conditioning has been demonstrated empirically, and the numerous instances in which such measures have been found to relate consistently and meaningfully to other variables provides strong justification for their use, irrespective of how they are interpreted. Over and above this pragmatic value, the utilization of awareness as a systematic concept has led to experimental findings which the Skinnerian approach would not have predicted and can explain only with piecemeal augmenting assumptions (Dulany, 1962; Spielberger, 1962).

In this paper we will contend that the descriptive behaviorists' rejection, on epistemological grounds, of verbal reports as indexes of mediating states has retarded the convergence of empirical findings and the development of adequate theory in verbal conditioning. We will argue moreover that this general approach has contributed little to the understanding of verbal behavior because, as Guttman (1963, p. 119) has recently suggested with respect to perception:

> . . . the effort to speak the language of behavior *only* is just inefficient, and to do so leads to bad experiments . . . , to a truncated set of laws of behavior, and to wide misinterpretation as to what our current knowledge of behavioral laws portends, theoretically and practically.

Having stated these goals and claims, we may now proceed to a discussion of procedures used for assessing awareness in verbal conditioning.

II. THE USE OF VERBAL REPORTS IN VERBAL CONDITIONING

The questioning techniques employed to assess awareness in verbal conditioning studies have included both oral interviews and written-response questionnaires. These have been utilized both during and after conditioning trials. In this section, we will first consider methodological factors associated with the assessment of awareness from postconditioning interviews. We will then describe procedures developed for assessing awareness during conditioning and the findings of two experiments in which such procedures were employed. Finally, the noncognitive formulations developed in Section I will be reviewed in the light of these studies.

The Assessment of Awareness after Conditioning

The assessment of awareness from postconditioning interviews is influenced by the style of questioning, the wording of questions and the number of questions asked, the interpolation of extinction trials, and the experimenter's concern with correlated hypotheses. Each of these factors will be discussed below in terms of its potential effect on reports of awareness obtained in postconditioning interviews.

Oral interview versus written questionnaire procedures. Both oral interviews and written-response questionnaires have been utilized for assessing awareness after the completion of conditioning trials. Either style of questioning would appear to have certain characteristic advantages and disadvantages. Written-response questionnaires eliminate the potential biasing influence of the experimenter who conducts an oral interview. However, the responses elicited by this procedure may not be as complete as those obtained from questions asked and answered orally, and clarification of unclear replies through additional inquiry is not generally feasible. Furthermore, some subjects may be inclined to avoid being incorrect by not responding when they are unsure of their hypotheses.

Number and wording of questions. In early verbal conditioning studies, awareness interviews were very brief, and the questions were vaguely worded. For example, in one study (Sidowski, 1954) the postconditioning interview consisted of only two questions: (*a*) "Were you aware of the purpose of the experiment?" and (*b*) "Were you aware of the purpose of the light?" (the contingent reinforcing stimulus). In response to these questions, a subject who might have been perfectly aware of the response-reinforcement contingency might not have had any idea of the "scientific purpose" of the experiment nor of the reinforcing stimulus. Furthermore, questions asked in this form make it easier to say "No" than "Yes" since the latter response will obviously require elaboration. A series of questions in which subjects were asked to state their thoughts about when and why the light was blinking would have been more satisfactory.

Given minimal questioning, subjects tend to respond according to their own preconceptions about what is important and refer to the reinforcing stimulus only when asked about it. For example, after the conclusion of a pseudo-ESP verbal conditioning experiment, Krieckhaus and Eriksen (1960) held informal conversations with subjects who showed performance gains, but failed to verbalize a correct response-reinforcement contingency in a structured interview. They reported:

> Two (subjects) verbalized very clearly their awareness of the effects of reinforcement upon their behavior. When asked why they had not stated this in response to the post-conditioning interview questions, both of them commented that they hadn't understood that this was what *E* was asking. They both stated that they thought *E* was concerned with whether or not they were getting ESP messages, rather than trying to find out whether the reinforcement of "good" had affected their choice of responses [p. 515].

Thus, the requirement of asking for information with clear and adequately worded questions is coordinate with the necessity of asking a sufficient number of questions.

Although more detailed questioning might be thought to increase the likelihood of suggesting awareness indiscriminately (Farber, 1963), a series of questions of gradually increasing specificity enables the interviewer to assess awareness and motivation while providing minimal *surplus* information. Furthermore, in studies in which such interviews were employed (see Spielberger, 1962), it has been demonstrated that: (*a*) Substantial numbers of subjects were identified as aware of correct response-reinforcement contingencies who were not so classified on the basis of brief interviews; and (*b*) performance gains were limited essentially to aware subjects and specific to those responses for which the individual subject was aware of a correct contingency.

Extinction trials. In verbal conditioning, the practice of interpolating extinction trials (e.g., Cohen, Kalish, Thurston, & Cohen, 1954; Greenspoon, 1955) or other tasks (Gergen, 1965; Sassenrath, 1959) between acquisition trials and the interview in which awareness is assessed probably reduces the likelihood that awareness will be detected. Subjects who are aware of a correct contingency during conditioning may either forget their hypotheses, interpret the withdrawal of the reinforcing stimulus during extinction as a sign that they were wrong, or both. We have observed that, even during acquisition trials, instances of *accidental* nonreinforcement due to experimenter error sometimes suffice to disconfirm subjects' hypotheses (Spielberger & DeNike, 1963).

Correlated hypotheses. The problem of assessing correlated hypotheses in learning experiments was first raised by Postman and Jarrett (1952). In verbal conditioning, a subject is considered to have a correlated hypothesis if he verbalizes a contingency which is different from, but correlated with, the reinforcement principle employed by the experimenter (Dulany, 1961). Thus, a subject need not be aware of the *experimenter's* definition of the contingency in order to show performance gains through conscious pursuit of his *own* hypotheses. For example, suppose one spring a whimsical millionaire were to emplace a few gold nuggets on some fertile ground on his estate and announce magnanimously to the townsfolk that all were welcome to come there and shovel around on the surface. Anyone inquiring among these townsfolk as they labored would find them highly aware that they were "digging for gold" and highly unaware that they were "spading a garden."

In the traditional verbal conditioning task in which plural nouns are reinforced (Greenspoon, 1955), a subject would receive 100% reinforcement if he said serially, "apples, oranges, peaches," etc. When interviewed, he might report as his hypothesis that reinforcement is contingent upon his naming "fruits." While the response class "fruits" is formally different from "plural nouns," the subject who named fruits would be reinforced for acting on this hypothesis and would be likely to develop the conviction that he was being encouraged to say words of similar meaning, as Dulany (1961) has convincingly shown. That subjects sometimes develop and pursue correlated ideas in response to less than 100% reinforcement points up the importance of *random* reinforcement rather than *no* reinforcement for control groups in verbal conditioning experiments (DeNike & Spielberger, 1963; Spielberger, DeNike, & Stein, 1965). And as Adams (1957) has pointed out, correlated hypotheses are typically *not* evaluated in investigations which have reported evidence for learning without awareness.

Thus far, we have discussed methodological factors which influence reports of awareness obtained from interviews or questionnaires conducted after the completion of the conditioning task. As previously noted, in recent verbal conditioning experiments in which such factors were taken into account, performance gains were limited essentially to subjects who verbalized a correct or correlated contingency. However, in these experiments, awareness was inferred from responses to interview questions asked *after* performance measures had been taken. Therefore, it could be argued that the performance gains of aware subjects were automatically produced by reinforcement and that the awareness-performance relationship could be accounted for by any of the noncognitive explanations described in Section I. The assessment of awareness during conditioning would provide a firmer basis for evaluating the temporal relationship between performance and awareness in verbal conditioning. Such procedures will now be considered.

The Assessment of Awareness during Conditioning

Most noncognitive explanations of verbal conditioning implicitly assume that increments in performance initially result from the direct and automatic effects of reinforcement and that performance gains occur prior to the time that the subject becomes aware of a correct response-reinforcement contingency. On the other hand, it is argued in cognitive explanations of verbal conditioning that awareness precedes performance increments. Thus, reinforcement theory and cognitive theory lead to differential predictions with regard to the temporal relationship between the development of awareness and the inception of performance gains. But in order to evaluate temporal factors in verbal conditioning, a procedure is required which will permit determination of whether or not subjects become aware of a correct contingency *during conditioning,* and if so, when. The results of two studies in which awareness was assessed during conditioning are discussed below.

Experiment I. DeNike (1964) reinforced female college students for giving *human-noun* responses in a word-naming task (Matarazzo, Saslow, & Pareis, 1960). His subjects were required to write down their "thoughts about the experiment" after each block of 25 response words. As a signal for the subject to record her thoughts, a light was turned on which remained lit until the subject indicated she was ready to resume

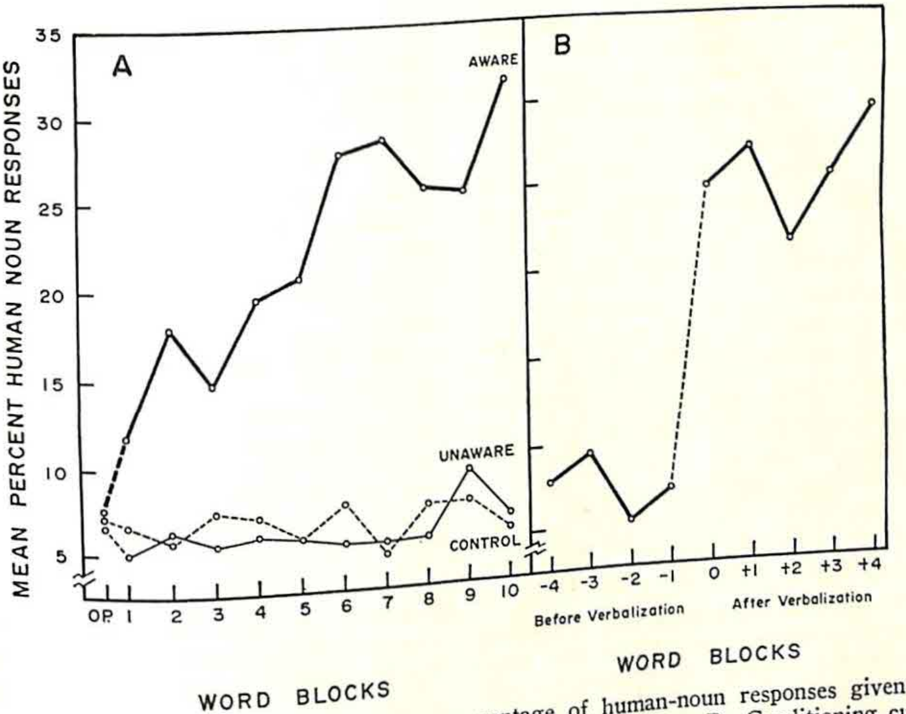

FIG. 2. Conditioning curves. (A: Mean percentage of human-noun responses given by the Aware, Unaware, and Control Groups on the conditioning task. B: Conditioning curve for the Aware Group in which the data for preverbalization and postverbalization word blocks have been separately Vincentized. Adapted from DeNike, 1964.)

saying words. No reinforcement was given during the first two word blocks, which provided a measure of operant rate. Beginning with the third word block, subjects in the Experimental Group were reinforced with "Mmm-hmm" for each human-noun response; those in the Control Group were reinforced with "Mmm-hmm" for 10% of their response words, according to a predetermined random schedule.

Awareness of the contingency between human-noun responses and the experimenter's "Mmm-hmm" was inferred from the thoughts (notes) which each subject recorded during conditioning. On the basis of these notes, subjects were independently rated by four judges as either *aware* or *unaware* of a correct contingency. Agreement between pairs of judges ranged from 90–95%, and there was unanimity or a 3–1 consensus among the judges with respect to the classification of 59 of the 61 subjects in the Experimental Group. For each aware subject, the judges also indicated the word block on which a correct contingency was first recorded.

If performance gains in verbal conditioning are automatically produced by reinforcement, it would be expected that gradual increments in performance would occur for all subjects, and that performance gains for aware subjects would occur prior to the word block on which they recorded a correct contingency in their notes. However, if acquisition of the reinforced response class in verbal conditioning is mediated by awareness, performance gains would be expected only for aware subjects, and these should occur on or subsequent to the word block on which such subjects recorded a correct contingency. The performance curves of DeNike's Aware, Unaware, and Control Groups are presented in Figure 2A. It may be noted that the output of human nouns for the Aware Group increased markedly over the reinforced word blocks and that the Unaware and Control Groups failed to show any performance gains. However, although only subjects who recorded correct contingencies in their notes gave more human nouns, it cannot be determined from the data as arrayed in Figure 2A whether or not aware subjects showed increments in performance prior to the word block on which they recorded a correct contingency.

In order to evaluate the temporal relation-

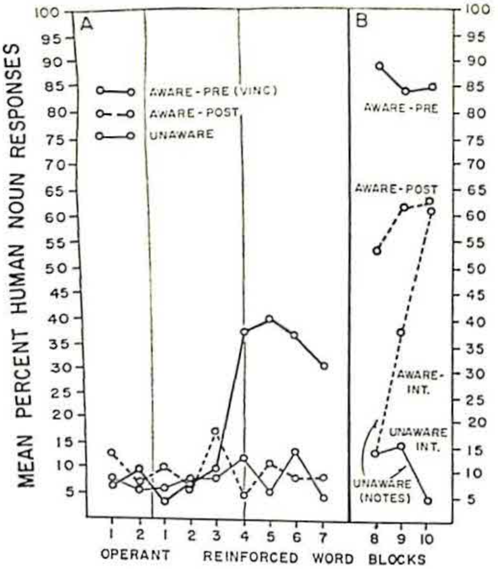

FIG. 3. Conditioning curves. (A: Mean percentage of human-noun responses given by the Aware-Pre, Aware-Post, and Unaware Groups on the operant word blocks and the reinforced word blocks prior to the incentive-inducing instruction. B: Mean percentage of human-noun responses given by the Aware-Pre, Aware-Post, Aware-Int., and Unaware-Int. Groups on the word blocks subsequent to the incentive-inducing instruction. Adapted from Spielberger et al., 1966.)

ship between performance gains and awareness, the conditioning data for the aware subjects were examined as a function of the word block on which each first recorded a correct contingency in her notes. This word block was designated the "0" block. Word blocks prior to and subsequent to Block 0 were designated the preverbalization and postverbalization blocks and were labeled respectively with negative and positive integers, after the practice of Philbrick and Postman (1955). The 0 blocks of the aware subjects were then aligned, and the data for the preverbalization and postverbalization blocks were separately Vincentized (Munn, 1950). The Vincentized conditioning curve for the Aware Group is given in Figure 2B in which it may be noted that performance on the conditioning task: (*a*) remained at essentially the same level in the preverbalization blocks as in the operant blocks, (*b*) increased markedly on the 0 block, and (*c*) was maintained at a relatively high level during the postverbalization blocks. The finding that the performance increments of aware subjects first occurred during the word block on which they first recorded a correct contingency in their notes would appear to indicate that their increased output of human-noun responses was cognitively mediated. The absence of preverbalization performance gains would appear particularly difficult to explain by the use of learning theories which ascribe automatic trans-situational reinforcing effects to verbal stimuli.

Although it was possible, according to a cognitive theoretical interpretation of verbal conditioning, for aware subjects in DeNike's study to give essentially 100% human-noun responses on each postawareness word block, it may be noted in Figure 2 that the mean number of such responses given by these subjects never exceeded 35%. Interview data obtained by DeNike after the conclusion of the conditioning task revealed some aware subjects for whom the reinforcing stimulus had no incentive value, and these subjects failed to show performance gains (DeNike, 1965). In contrast, aware subjects who reported that they wanted to receive reinforcement showed a marked increase in their output of human nouns on postawareness word blocks. In the next study to be reported, a uniformly high level of motivation to receive reinforcement was induced during conditioning in all subjects.

Experiment II. Using similar conditioning procedures to those described for Experiment I, female college students were reinforced with "Mmm-hmm" for giving human-noun responses (Spielberger, Bernstein, & Ratliff, 1966). Also, as in Experiment I, awareness of a correct response-reinforcement contingency was evaluated from the thoughts about the experiment (notes) which subjects recorded after each word block. However, the procedures differed from those of the previous experiment in that, between the seventh and eighth reinforced word blocks of the conditioning task, each subject was told: "As you may have noted, there is a rule under which I say 'Mmm-hmm.' Try to act on that rule so as to make me say 'Mmm-hmm' as often as you can." The purpose of this instruction was to increase the incentive value of the reinforcing stimulus. Immediately after conditioning a different experimenter conducted an interview which checked on the incentive manipulation and provided additional data on subjects' awareness.

The notes which subjects recorded during conditioning were rated by two judges who had neither contact with the subjects nor knowledge of their performance on the conditioning task. The judges agreed perfectly in classifying each subject as either aware or unaware of a correct contingency. For the subjects rated aware, the judges also indicated the word block on which a correct contingency was first reported. Those subjects who recorded a correct contingency prior to the incentive-inducing instruction were designated the Aware-Pre Group; those who did so subsequent to this instruction were designated the Aware-Post Group; subjects who failed to record a correct contingency during conditioning were called the Unaware Group. The mean percentage of human-noun responses given by the three groups in the operant blocks and the reinforced blocks prior to the incentive-inducing instruction are given in Figure 3A in which the Aware-Pre Group's conditioning data have been Vincentized. Statistical analyses of these data indicated that: (a) The three groups did not differ during the operant blocks; (*b*) the performance gains of the Aware-Pre Group first occurred on the word block on which these subjects first recorded a correct contingency; and (*c*) the Aware-Post and Unaware Groups failed to show performance gains. Thus, prior to the incentive-inducing instruction, aware and unaware subjects performed much as did their counterparts in DeNike's study (see Figure 2).

Data obtained in the postconditioning interview indicated that the instruction to try to make the experimenter say "Mmm-hmm" had the desired effect on the subjects' motivation to receive reinforcement. Almost all the subjects indicated that, as a consequence of the instruction, they wanted more for the experimenter to say "Mmm-hmm," and they tried harder to make him say it. The performances of the Aware-Pre, the Aware-Post, and Unaware Groups in the word blocks subsequent to the instruction are indicated in Figure 3B. The immediate marked increase in the number of human nouns given by the Aware-Pre Group on Block 8 suggested that the incentive manipulation increased the motivation of these subjects to receive reinforcement and that their prior awareness of a correct contingency permitted them at once to give more human-noun responses. The somewhat less marked increase in human nouns for the Aware-Post Group subsequent to the instruction may be interpreted as indicating that these subjects, most of whom recorded a correct contingency in their notes on Block 8, had to become aware before heightened motivation induced by the instruction could influence their performance. Consistent with this interpretation, examination of the conditioning data for individual subjects in the Aware-Post Group revealed that initial performance increments tended to occur on the word block on which a correct contingency was first recorded.

The mean percentage of human-noun responses given by the Unaware Group increased gradually in the postinstruction word blocks, from 15% on Block 8 to 30% on Block 10 (not shown in Figure 3B), suggesting that these subjects conditioned without awareness. In order to evaluate this possibility further, the postconditioning interview protocols of all subjects who failed to record a correct contingency in their notes were examined by two judges who had neither contact with the subjects nor knowledge of their performance on the conditioning task. The judges agreed perfectly in rating approximately half of these subjects as aware of a correct contingency solely on the basis of their verbal reports in the interview.

The performance of subjects rated aware on the basis of their interview responses, but unaware on the basis of their notes (Aware-Int. Group), is compared, in Figure 3B, with the performance of subjects rated unaware on the basis of *both* their interview responses and their notes (Unaware-Int. Group). For the latter group, the output of human-noun responses subsequent to the instruction did not increase. Indeed, by the final word block, the number of human nouns given by the Unaware-Int. Group was below what it had been prior to the instruction. In contrast, the conditioning curve for the Aware-Int. Group showed a significant rise subsequent to the instruction, and performance increments for individual subjects tended to correspond with the word block on which they claimed they became aware. Furthermore, several of these subjects spontaneously reported that they had not recorded a correct response-reinforcement contingency in their notes because they did not become aware of it until the final word block and did not have sufficient confidence in their hypothesis to report it.

In sum, the findings in Experiments I and II are consistent with the hypothesis that "what is learned" in verbal conditioning is awareness of a correct response-reinforcement contingency. These findings would also appear to indicate that the reinforcing stimulus in verbal conditioning has both information and incentive value (Dulany, 1962), and that the latter influences the degree to which subjects who are aware of a correct contingency act on their awareness. The adequacy with which noncognitive explanations of verbal conditioning can account for the results in these experiments is considered below.

Noncognitive Interpretations of the Findings in Experiments I and II

The findings in Experiments I and II would appear difficult to account for in terms of a theory which does not include a concept of awareness as a mediating cognitive state or process. In these investigations, only subjects who were judged to be aware of a correct response-reinforcement contingency showed performance gains, and these tended to correspond with the word block on which each aware subject first recorded a correct contingency. However, the question remains whether the obtained relationships between performance gains and operational indexes of awareness can be explained without recourse to cognitive variables. The possibility that these relationships arose artifactually will now be examined in terms of each of the noncognitive formulations described in Section I (see Figure 1).

Suggestion. The possibility that awareness might have been suggested to subjects who showed conditioning through cues provided by a biased experimenter was practically eliminated in Experiments I and II. In these studies, awareness was inferred primarily from notes recorded during conditioning, and these notes were *not* written in response to potentially biasing questions. The signal for the subject to record her thoughts, a light presented at uniform intervals, was a completely impersonal nonverbal stimulus which was in no way contingent upon the subject's conditioning performance. For those subjects in Experiment II whose awareness was inferred from their responses to the postconditioning interview, the experimenter who conducted this interview had neither prior contact with the subjects nor knowledge of their performance on the conditioning task.

Labeling. The findings in Experiments I and II do not support the interpretation that reinforcement automatically produced increments in performance and that the subjects' knowledge of a correct response-reinforcement contingency subsequently resulted from their noticing and labeling the relationship between the critical response class and the reinforcing stimulus. As may be noted in Figures 2B and 3A the performance gains of aware subjects first occurred on the word block on which these subjects recorded a correct contingency in their notes. Thus, the reinforcing stimulus had no apparent influence on performance prior to the subjects becoming aware of a correct contingency. And while it still might be argued that automatic reinforcement effects produced performance gains during the 0 word block, the absence of performance gains on the preceding word blocks would remain unaccounted for. Only the performance of the small group of subjects ($N = 4$) in the Aware-Int. Group of Experiment II might be interpreted as consistent with the labeling hypothesis. But even these subjects, who failed to record a correct contingency in their notes but reported it in the postconditioning interview, tended spontaneously to explain this discrepancy in terms of insufficient opportunity to test their hypotheses.

Secondary mediation. The secondary-mediation hypothesis assumes that automatic performance increments precede the development of awareness of a correct contingency, and that awareness then facilitates further performance gains. This view is contradicted by the same findings in Experiments I and II which led us to challenge the labeling interpretation. The demonstration of automatic reinforcement effects prior to the verbalization of awareness would appear crucial in order to support the hypothesis of secondary mediation. No such evidence was found, except possibly for the subjects in the Aware-Int. Group in Experiment II. Thus, at the very least, the secondary-mediation hypothesis would require extensive modification if it is to account for *all* of the findings in Experiments I and II.

Joint conditioning effects. According to the joint conditioning hypothesis, gradual gains in performance on the conditioning task would be expected along with the simultaneous strengthening of the covert "awareness response." The absence of performance increments in Experiments I and II prior to the word block on which aware subjects recorded a correct contingency, and the striking temporal correspondence of the inception of performance gains with the verbalization of awareness would seem to dictate the additional assumption of a "threshold" concept in order to support the view that the reinforced response class and a covert awareness response were simultaneously and automatically strengthened. Moreover, it would also appear necessary to assume that the threshold at which awareness was verbalized corresponds to the threshold at which the cumulative effects of reinforcement influence performance on the conditioning task, and that performance and covert awareness responses were practically identically susceptible to influence by reinforcement in *each individual subject*. While such assumptions might provide tenuous post-hoc explanations for the findings obtained in Experiments I and II, the conclusion appears inescapable that a theory of automatic strengthening of verbal response by reinforcement requires extensive modification and radical augmentation before it can account for such findings. In the next section, a theoretical analysis of verbal conditioning in terms of the concepts of cognitive learning theory will be proposed.

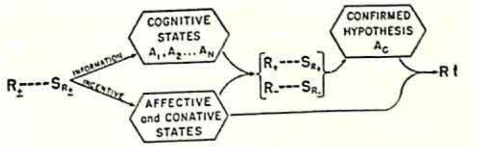
COGNITIVE STATES $A_1, A_2 \ldots A_N$
CONFIRMED HYPOTHESIS A_C
INFORMATION
INCENTIVE
$R_\pm$----$S_{R_\pm}$
R_+---S_{R_+}
R_----S_{R_-}
AFFECTIVE and CONATIVE STATES
R↑

FIG. 4. Schematic diagram indicating the hypothetical processes which are assumed to mediate performance gains in verbal conditioning for subjects who show "conditioning" effects.

III. AN ANALYSIS OF VERBAL CONDITIONING IN TERMS OF COGNITIVE CONCEPTS

The schematic diagram presented in Figure 4 reflects a general formulation of the sequence of hypothetical events which we believe mediate performance gains in verbal conditioning for subjects who show "conditioning" effects. The diagram is not intended to represent any particular cognitive theory in detail. Rather, it is offered to account for the findings of recent investigations of verbal conditioning, such as those described in Section II, at the same level of specificity as the noncognitive formulations presented in Figure 1. A similar more detailed and sophisticated theoretical analysis of verbal conditioning has been recently advanced by Dulany (1962).

The temporal sequence of events in Figure 4 is from left to right. Those events which are merely temporally contiguous are indicated by dotted lines; presumed causal relationships between events are represented by solid arrows. The symbol $R_\pm$ indicates that in verbal conditioning the subject initially emits a variety of responses, some of which belong to the reinforced class, R_+, and some of which do not, R_-. These are followed respectively by reinforcement, S_{R+}, and nonreinforcement, S_{R-}, which convey differential information and provide differential incentive to the subject.

The information provided by the reinforcing stimulus gives rise to one or more cognitive states, represented as $A_1, A_2, \ldots A_N$. For example, the sequence of cognitive states developed by a subject in a word-naming verbal conditioning task might be: "The experimenter sometimes says 'good' after some of my words" (A_1); "Perhaps his saying 'good' depends on what words I say" (A_2); "Perhaps he wants me to figure out what words he is saying 'good' to, and give more of those" (A_3), etc. Depending upon the pattern and amount of reinforcement which they receive and recall, some subjects may be able to skip some of these early logical steps and go directly to a tentative formulation of the correct hypothesis, A_N: "Perhaps he is saying 'good' when I say words denoting people." We have observed that subjects who verbalize a correct hypothesis approach the experiment essentially as a problem-solving task. Furthermore, in verbal conditioning experiments in which subjects are given explicitly a problem-solving set, they proceed more directly to test out their hypotheses, and a larger proportion verbalize a correct contingency (DeNike & Spielberger, 1963; Spielberger & Levin, 1962).

Concurrent with the concept-formation activity involved in arriving at a correct hypothesis, subjects also develop affective and conative states based on the *incentive* provided by the reinforcing stimulus. For example, if the reinforcing stimulus has positive incentive value for a particular subject, he will *want* to receive reinforcement, and generally will *try* to elicit it from the experimenter. On the other hand, if the reinforcing stimulus has neutral or negative incentive value for the subject, he may be indifferent to it or attempt to avoid it (Mandler & Kaplan, 1956). Thus, the subject's affective and conative states in conjunction with his cognitive states (hypotheses) lead him to give responses, R_+, which he believes are in the reinforced response class, intermixed with a few responses, R_-, which are not expected to elicit reinforcement. Following these responses, the selective administration of the reinforcing stimulus, S_{R+}, or the absence of the reinforcing stimulus, S_{R-}, either confirms the subject's hypothesis, A_C, or leads him to modify it. The confirmed hypothesis, if accompanied by appropriate affective and conative states, results in increased output of the reinforced response class, $R\uparrow$.

It should be noted that the schematic diagram in Figure 4 massively oversimplifies our view of verbal conditioning. In it, we have treated affective and conative states as essentially constant and independent of cognitive states; there is much reason to suspect that these fluctuate and interact with the subjects' hypotheses. We have diagramed the process for the successful subject and ignored the numerous blind alleys explored by both successful and unsuccessful subjects; such processes should receive consideration in any theory which claims completeness. The nature of the conditioning task, the status of the experimenter, the instructions given to the subject, the characteristics of the subject population—all of these have demonstrable influence on mediating states and performance in verbal conditioning. In the final section of this paper, the significance of the verbal conditioning paradigm for the more general question of human learning without awareness will be considered.

IV. VERBAL CONDITIONING AND THE QUESTION OF LEARNING WITHOUT AWARENESS

Few persons doubt that behavior occurs without awareness, as in, for example, shifting gears, fingernail biting, facial tics, etc. (Eriksen, 1960). Of theoretical interest, however, is the extent to which such behaviors are initially *learned* without the mediation of cognitive processes. Kimble (1962) has suggested that the question of whether learning without awareness is possible may not have the same answer for all forms of learning. In classical eyelid conditioning, according to Kimble, it would appear that the subject need only be aware of the unconditioned stimulus for learning to take place.

In verbal conditioning, the evidence for learning without awareness reported in early experiments (e.g., Cohen et al., 1954; Greenspoon, 1955; Taffel, 1955) now appears suspect because of shortcomings in the evaluation of awareness in these studies. In recent experiments in which awareness was more carefully and thoroughly assessed (e.g., DeNike, 1964; Dulany, 1961, 1962; Spielberger, 1962), the absence of learning-without-awareness effects all but compels the conclusion that performance gains in verbal conditioning are consciously mediated.[6] However, for the reasons indicated below, the verbal conditioning paradigm may not provide an appropriate framework for clarifying the more general question of whether human operant learning can occur without awareness.

Subject Vigilance

Most subjects in psychological experiments, as Orne (1962) has pointed out, tend to be alert to manipulations that the experimenter may impose upon them and vigilant vis-à-vis their own behavior. In almost all verbal conditioning research, the subject is aware that he is participating in a psychological experiment. Thus, subjects in verbal conditioning experiments ordinarily notice the reinforcing stimulus and attempt to ascribe some meaning to it, making learning without awareness unlikely.

Response Monitoring

Despite reinforcement and no admonition to the contrary, reinforced words are rarely repeated in word-naming verbal conditioning tasks. This almost universal tendency not to repeat words would be difficult to explain by automatic reinforcement theories. When asked why they did not repeat reinforced words, subjects usually indicate that they believed they were supposed to give a variety of words. It would appear that the subject keeps both his past responses and their consequences in mind and monitors new responses according to his hypotheses about the experiment. Thus, on the assumption that selective responding is guided by response monitoring in accordance with the subjects' hypotheses, the tendency not to repeat reinforced words in verbal conditioning becomes understandable, and learning without awareness seems improbable.

---

[6] It should be noted, however, that some subjects who are unaware of a correct contingency may nevertheless show slight gains in performance in verbal conditioning if they develop what may be termed a *correlated set*. For example, subjects (patients) who regard the Taffel (1955) sentence-construction task as a personality test sometimes assume they are supposed to make up sentences about themselves and their families. In response to either *random or systematic* reinforcement, subjects with such response dispositions may interpret the reinforcing stimulus as a sign of the experimenter's approval or encouragement and therefore give increasingly more sentences beginning with "I" or "We" (Spielberger, DeNike, & Stein, 1965).

Response Sets

Some subjects approach verbal conditioning with a problem-solving set, while others interpret the conditioning task as either a personality test or an attempt to make them "conform." When a problem-solving set is combined with response monitoring in a vigilant subject, the subject notices the reinforcing stimulus, proceeds to test out successive hypotheses about the experiment, and is likely to become aware of a correct or correlated response-reinforcement contingency. In contrast, subjects with a personality-test set may get so involved in their free associations (on a word-naming verbal conditioning task) that they are oblivious to the reinforcing stimulus or tend to regard it as annoying and disruptive. Other subjects with personality-test sets devote considerable effort to the selection of "innocuous" or "unrevealing" words or otherwise defensively edit their responses so that they will not reveal personal secrets or shortcomings. Subjects who believe the experimenter is attempting to manipulate their behavior often report that they deliberately ignored the reinforcing stimulus or tried to avoid it. Thus, while a "conformity" or "personality" interpretation of the verbal conditioning task tends to prevent learning without awareness by causing subjects to ignore or avoid the reinforcing stimulus and/or to perceive it as unpleasant, a problem-solving approach tends to preclude learning without awareness by making it easier for the subjects to become aware.

The foregoing analysis of the "demand characteristics" (Orne, 1962) of verbal operant conditioning suggests that it is unlikely that learning without awareness will be found in laboratory experiments which employ the methods traditionally associated with the verbal conditioning paradigm. Under such conditions, it might be expected on a priori grounds that subjects will be particularly vigilant about their own verbal responses which are likely to be more carefully monitored than most other behavior. Or, in Eriksen's (1960) words, ". . . any situation where the cues and the reinforcements are salient enough to produce learning, will not escape detection by awareness [p. 298]."

It would appear more plausible to seek evidence of learning without awareness in situations where subjects are not aware that they are participating in a psychological experiment and on tasks which do not induce intensive response monitoring or defensive response sets. A likely place to look for such evidence might be in settings where subjects are motivated not to be aware, as, for example, in experimental situations involving processes such as repression, dissonance reduction, conformity, and ingratiation. At present, only scattered attempts have been made to investigate human operant learning with research designs representative of some of the forenamed conditions (e.g., Centers, 1963; Gergen, 1965; Goldiamond, 1965; Hefferline, Keenan, & Harford, 1959; Kimmel & Baxter, 1964; Verplanck, 1956; Vogel-Sprott, 1964). In such studies, however, the role of awareness has ordinarily not been the focus of investigation, and, consequently, the methods employed to assess awareness are subject to the criticisms discussed earlier. Therefore, where learning-without-awareness effects have been found, little confidence may be placed in such results. This in no way implies that human operant learning cannot take place without awareness, but merely affirms

Adams' (1957) observation that such learning has not been demonstrated convincingly in the laboratory.

Some Final Considerations

We have endeavored in this paper to compare the relative adequacy with which descriptive behaviorism and cognitive learning theory can account for the findings of verbal operant conditioning experiments. We have argued that the descriptive behaviorists' implicit rejection of awareness as a concept has had serious methodological consequences which have retarded the convergence of empirical findings and the development of adequate theory. We have also contended that the utilization of awareness as a systematic concept has led to experimental findings which the Skinnerian approach would not have predicted. Finally, we have presented evidence to support the view that the reinforcing stimulus in verbal conditioning has both information and incentive value and that "what is learned" in verbal conditioning is awareness of a correct (or correlated) response-reinforcement contingency.

With respect to whether human operant learning can occur without awareness, it should be noted that although this question would appear to have an empirical answer, in verbal conditioning the answer turns out to be inextricably tied to an investigator's theoretical orientation and, more fundamentally, to his epistemological assumptions concerning awareness as a concept (Spielberger, 1962, 1965). The issue is further complicated by the fact that different learning theories imply different operational definitions of the learned response (Campbell, 1954), leading in verbal conditioning to the collection of noncomparable data and to fruitless theoretical controversy. This controversy, which revolves about the role of awareness in verbal conditioning, is reminiscent of earlier disagreements among S–R and cognitive theorists concerning the role of reinforcement in learning. It has since been recognized that much of the controversy and confusion with respect to reinforcement stemmed from the failure of learning theorists to differentiate conceptually between the law of effect as an empirical statement and as a general theory of reinforcement (Spence, 1951).

On an empirical level, most would agree with the observation that performance in a variety of tasks is facilitated by reinforcement, but S–R and cognitive learning theories are still sharply divided on the question of whether reinforcement is required for learning to take place. Similarly, it is generally accepted that performance will be facilitated by the presence of appropriate cognitive states; that is, an "empirical law of cognition" is as supportable as the empirical law of effect,[7] but whether cognitive processes mediate any and all forms of learning is as dubious as a general theory of reinforcement. Thus, while a "theoretical law of cognition" might well serve to explain the findings in verbal conditioning experiments with human adults, it is quite another matter to demonstrate the role of cognitive processes in the learning of animals and preverbal children.

Nevertheless, the careful investigation of subjects' awareness in relation to other behavior is, we believe, requisite to the development of an adequate theory of human learning. Progress in this direction will depend upon the general acceptance of experimental procedures which permit sensitive evaluation of cognitive processes. These procedures must surely include subjects' verbal reports and must provide necessary safeguards against biasing or distorting such reports. Furthermore, it must be clearly recognized that the validity of verbal reports in any experimental context rests upon the willingness and linguistic competence of subjects to describe their mediating states when properly questioned (Dulany, 1961), and that verbal reports are but imperfectly related to the subjects' internal states (Eriksen, 1960). Paradoxically, perhaps the most significant contribution of verbal conditioning research has been the stimulation of interest in verbal report procedures and in concepts such as awareness among psychologists who are inclined to insist that thoughts and ideas are beyond the limits of scientific inquiry.

---

[7] We are grateful to Donald L. Thistlethwaite for suggesting the analogy between an empirical law of cognition in verbal conditioning and the empirical law of effect in traditional discussions of learning theory.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychological Bulletin,* 1957, **54**, 383–405.

CAMPBELL, D. T. Operational delineation of "what is learned" via the transposition experiment. *Psychological Review,* 1954, **61**, 167–174.

CAMPBELL, D. T. Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 6. New York: McGraw-Hill, 1963. Pp. 94–172.

CENTERS, R. A. Laboratory adaptation of the conversational procedure for the conditioning of verbal operants. *Journal of Abnormal and Social Psychology,* 1963, **67**, 334–339.

COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *Journal of Experimental Psychology,* 1954, **47**, 106–110.

DENIKE, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology,* 1964, **68**, 521–529.

DENIKE, L. D. Recall of reinforcement and conative activity in verbal conditioning. *Psychological Reports,* 1965, **16**, 345–346.

DENIKE, L. D., & SPIELBERGER, C. D. Induced mediating states in verbal conditioning. *Journal of Verbal Learning and Verbal Behavior,* 1963, **1**, 339–345.

DULANY, D. E. Hypotheses and habits in verbal "operant conditioning." *Journal of Abnormal and Social Psychology,* 1961, **63**, 251–263.

DULANY, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962. Pp. 102–129.

DULANY, D. E. How can we speak of awareness and volition as instrumental? In H. D. Kimmel (Chm.), Awareness as a factor in verbal operant conditioning. Symposium presented at Southeastern Psychological Association, Miami Beach, April 1963.

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: University of Nebraska Press, 1958. Pp. 169–228.

ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review,* 1960, **67**, 279–300.

ERIKSEN, C. W. Figments, fantasies, and follies: A search for the subconscious mind. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962. Pp. 3–26.

FARBER, I. E. The things people say to themselves. *American Psychologist,* 1963, **18**, 185–197.

GERGEN, K. J. The effects of interaction goals and personalistic feedback on the presentation of self. *Journal of Personality and Social Psychology,* 1965, **1**, 413–424.

GOLDIAMOND, I. Stuttering and fluency as manipulable operant response classes. In L. Krasner & L. P. Ullmann (Eds.), *Research in behavior modification: New developments and their clinical implications.* New York: Holt, Rinehart, & Winston, 1965. Pp. 106–156.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *American Journal of Psychology,* 1955, **68**, 409–416.

GREENSPOON, J. Verbal conditioning and clinical psychology. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology.* New York: Basic Books, 1962. Pp. 510–553.

Greenspoon, J. Reply to Spielberger and DeNike: "Operant conditioning of plural nouns: A failure to replicate the Greenspoon effect." *Psychological Reports*, 1963, **12**, 29–30.

Guttman, N. Laws of behavior and facts of perception. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 5. New York: McGraw-Hill, 1963. Pp. 114–178.

Hefferline, R. F., Keenan, B., & Harford, R. A. Escape and avoidance conditioning in human subjects without their observation of the response. *Science*, 1959, **130**, 1338–1339.

Irwin, F. W., Kauffman, K., Prior, G., & Weaver, H. B. On "learning without awareness of what is being learned." *Journal of Experimental Psychology*, 1934, **17**, 823–827.

Kanfer, F. H., & Marston, A. R. Verbal conditioning, ambiguity and psychotherapy. *Psychological Reports*, 1961, **9**, 461–475.

Kanfer, F. H., & Marston, A. R. The effect of task-relevant information on verbal conditioning. *Journal of Psychology*, 1962, **53**, 29–36.

Kanfer, F. H., & McBrearty, J. F. Verbal conditioning: Discrimination and awareness. *Journal of Psychology*, 1961, **52**, 115–124.

Kimble, G. A. Classical conditioning and the problem of awareness. In C. W. Eriksen (Ed.), *Behavior and awareness*. Durham: Duke University Press, 1962. Pp. 27–45.

Kimmel, H. D., & Baxter, R. Avoidance conditioning of the GSR. *Journal of Experimental Psychology*, 1964, **68**, 482–485.

Krasner, L. The therapist as a social reinforcement machine. In H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy*. Vol. 2. Washington, D. C.: American Psychological Association, 1962. Pp. 61–94.

Krasner, L., & Ullmann, L. P. Variables affecting report of awareness in verbal conditioning. *Journal of Psychology*, 1963, **56**, 193–202.

Krieckhaus, E. E., & Eriksen, C. W. A study of awareness and its effect on learning and generalization. *Journal of Personality*, 1960, **28**, 503–517.

Mandler, G., & Kaplan, W. K. Subjective evaluation and reinforcing effect of a verbal stimulus. *Science*, 1956, **124**, 582–583.

Matarazzo, J. D., Saslow, G., & Pareis, E. N. Verbal conditioning of two response classes: Some methodological considerations. *Journal of Abnormal and Social Psychology*, 1960, **61**, 190–206.

Munn, N. L. *Handbook of psychological research on the rat: An introduction to animal psychology*. Boston: Houghton Mifflin, 1950.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776–783.

Philbrick, E. B., & Postman, L. A further analysis of "learning without awareness." *American Journal of Psychology*, 1955, **68**, 417–424.

Postman, L., & Sassenrath, J. M. The mental analysis of learning without awareness. *American Journal of Psychology*, 1952, **65**, 244–255.

Postman, L., & Sassenrath, J. M. The automatic action of verbal rewards and punishments. *Journal of General Psychology*, 1961, **65**, 109–136.

Rosenthal, R. On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Scientist*, 1963, **51**, 268–283.

Salzinger, K. Experimental manipulation of verbal behavior: A review. *Journal of General Psychology*, 1959, **61**, 65–94.

Sassenrath, J. M. Learning without awareness and transfer of learning sets. *Journal of Educational Psychology*, 1959, **50**, 205–212.

Sidowski, J. B. Influence of awareness of reinforcement on verbal conditioning. *Journal of Experimental Psychology*, 1954, **48**, 355–360.

Skinner, B. F. *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts, 1938.

Skinner, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. Behaviorism at fifty. *Science*, 1963, **140**, 951–958. (a)

Skinner, B. F. Operant behavior. *American Psychologist*, 1963, **18**, 503–515. (b)

Spence, K. W. The postulates and methods of "behaviorism." *Psychological Review*, 1948, **55**, 67–78.

Spence, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 690–729.

Spielberger, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behaviorism and awareness*. Durham: Duke University Press, 1962. Pp. 73–101.

SPIELBERGER, C. D. Theoretical and epistemological issues in verbal conditioning. In S. Rosenberg (Ed.), *Directions in psycholinguistics.* New York: Macmillan, 1965. Pp. 149–200.

SPIELBERGER, C. D., BERNSTEIN, I. H., & RATLIFF, R. G. Information and incentive value of the reinforcing stimulus in verbal conditioning. *Journal of Experimental Psychology,* 1966, **71**, 26–31.

SPIELBERGER, C. D., & DENIKE, L. D. Implicit epistemological bias and the problem of awareness in verbal conditioning: A reply to Greenspoon. *Psychological Reports,* 1963, **12**, 103–106.

SPIELBERGER, C. D., DENIKE, L. D., & STEIN, L. S. Anxiety and verbal conditioning. *Journal of Personality and Social Psychology,* 1965, **1**, 229–239.

SPIELBERGER, C. D., & LEVIN, S. M. What is learned in verbal conditioning? *Journal of Verbal Learning and Verbal Behavior,* 1962, **1**, 125–132.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology,* 1955, **51**, 496–501.

THORNDIKE, E. L., & ROCK, R. T. Learning without awareness of what is being learned or intent to learn it. *Journal of Experimental Psychology,* 1934, **17**, 1–19.

VERPLANCK, W. S. The operant conditioning of human motor behavior. *Psychological Bulletin,* 1956, **53**, 70–83.

VOGEL-SPROTT, M. E. Response generalization under verbal conditioning in alcoholics, delinquents, and students. *Behaviour Research and Therapy,* 1964, **2**, 135–141.

(Received March 22, 1965)

Psychological Review
1966, Vol. 73, No. 4, 327–334

CRITERIA OF EXPECTANCY[1]

FRANCIS W. IRWIN

University of Pennsylvania

Expectancies of the form "aa'Eo" (i.e., "Act a rather than Act a' is expected to lead to Outcome o") are diagnosed by positive association between the relative frequencies of a pair of alternative acts, a and a', and a pair of opposed preference states, oPo' and o'Po. Controls to exclude artifactual association due to response biases are described. Properties of such expectancies and experimental problems that arise in diagnosing them are discussed, and comparisons with earlier formalizations are made.

Although the concept of expectancy is central to "cognitive" theories of behavior, proponents of such theories may be said to have had more success in using it polemically than in analyzing it and setting forth its empirical criteria. What follows is an attempt in the latter direction.

EMPIRICAL CRITERIA OF ACT-OUTCOME EXPECTANCIES

In simplest terms, the fundamental notion of cognitive theory is that an organism's choices are determined by what it wants and by what it expects. If so, then to know what it expects, we must know how its choices vary with what it wants. To illustrate:

1. Suppose that a child confronted with two boxes opens the blue one when he prefers a toy to a trinket and opens the yellow one when he prefers a trinket to a toy. Then a cognitive theorist will conclude that the child expects that opening the blue box rather than the yellow one will get him a toy and that opening the yellow box rather than the blue one will get him a trinket.

2. If a rat in a T maze turns right when it prefers food to water and turns left when it prefers water to food, then according to cognitive theory it expects turning right rather than turning left to lead to food and turning left rather than turning right to lead to water.

The paradigm appropriate to testing for such act-outcome expectancies is represented in Table 1. Let a and a' be any pair of alternative acts of which one and only one occurs in any trial,

TABLE 1

PARADIGM FOR ESTABLISHING EXISTENCE OF THE EXPECTANCIES $aa'Eo$, $a'aEo'$, $aa'Eo'$, AND $a'aEo$

Alternative acts	Conditions			
	1 oPo'	2 $o'Po$	1' oPo'	2' $o'Po$
a	+ (o)	− (o)	− (o')	+ (o')
a'	− (o')	+ (o')	+ (o)	− (o)

Note.—o, o': differential outcomes; oPo', $o'Po$: opposed preference states. Plus and minus signs indicate association between acts and preference states.

---

[1] This investigation was supported in part by Research Grant MH-06580 to the University of Pennsylvania from the National Institute of Mental Health, United States Public Health Service. The author is grateful to R. L. Solomon, R. Duncan Luce, and Joan Gay Snodgrass for valuable criticism. The following terms will be used in the senses given them in previous papers (Irwin, 1958, 1961): act, outcome, differential outcomes, null differential outcome (ø), preference, discrimination.

and let *o* and *o′* be any pair of differential outcomes (Irwin, 1961). In Condition 1, *o* is preferred to *o′* (*oPo′*); in Condition 2 the reverse obtains (*o′Po*). Then a positive association between the relative frequencies of Acts *a* and *a′* and the preference states *oPo′* and *o′Po,* as suggested by the plus and minus signs in the cells of the table, is evidence for the existence of two expectancies, namely, that Act *a* rather than Act *a′* will lead to Outcome *o,* and that Act *a′* rather than Act *a* will lead to Outcome *o′*. These two expectancies may be symbolized by *aa′Eo* and *a′aEo′*, respectively, and *o* and *o′* may be called their expectanda.

It is possible, however, for such results to reflect merely an artifact of response bias. For example, if a bias favoring *a* over *a′* is for some reason associated with *oPo′* but not with *o′Po,* then Conditions 1 and 2 will show an association between the acts and the preference states, but inferences concerning expectancies will be unwarranted. The experimenter may know his organism well enough not to be worried by this possibility, but to eliminate it experimentally requires only the control procedures of Conditions 1′ and 2′, which attempt to reverse the direction of the association found in Conditions 1 and 2. If this is successful, not only is the artifact ruled out, but the existence during the control trials of two different expectancies, *aa′Eo′* and *a′aEo,* is demonstrated. Such control procedures are often carried out as a matter of course, although sometimes with subjects (*S*s) other than those of Conditions 1 and 2 (see, for example, Kendler, 1946).

Although Table 1 constitutes the paradigm for establishing whether or not certain expectancies exist at the time of testing, it implies nothing about how they may have been acquired. One could accept from a colleague a group of *S*s—rats, children, monkeys, adults—and, in complete ignorance of their past, still undertake to find out whether they had some specified expectancy. This expectancy might have been induced by training *S*s with repeated trials in a design like that of Conditions 1 and 2 of Table 1, the outcome of each trial being what is represented in parentheses in the cells of the table. Or, in the case of organisms capable of language, *S*s might have been told what to expect as the result of their choices in the situation. Or *S*s might have been exposed to the sight of other *S*s making choices in the situation and obtaining the outcomes from them, as in experiments on "imitation." What is needed is not information about this history but rather control over *S*s' preferences; with this, the expectancies can be diagnosed if they exist.

COMMENTS AND DISCUSSION

Formal Properties

The following properties of the expectancies under discussion will be postulated.

1. Since *a* and *a′* are taken to be alternative acts in a binary-choice situation, they are nonidentical and mutually exclusive. Hence, the expression *aaEo* is not defined.

2. At any moment in time, with *o* held constant: If *aa′Eo,* then not *a′aEo;* that is, *aa′Eo* is asymmetrical in *a* and *a′*. But (not *aa′Eo*) does not imply *a′aEo*.

3. At any moment, with *o* held constant: If *aa′Eo* and *a′a″Eo,* then *aa″Eo;* that is, the relation between the acts is transitive.

In discussing the critical experiment required for establishing that an expectancy of the type *aa′Eo* exists, it was pointed out that obtaining the evidence for one such expectancy entails obtaining evidence also for another expectancy closely related to the first. This linkage can be expressed symbolically:

$$(a)(a')\quad (\exists o)aa'Eo \equiv (\exists o')a'aEo'. \qquad [1]$$

That is to say: Given any pair of alternative acts, to assert that an organism expects one act rather than the other to lead to some outcome, o, is materially equivalent to asserting that there is some other outcome, o', such that the organism expects the second act rather than the first to lead to it.

The freedom to choose o' in Formula 1 is material. For example, one of the expectancies in Illustration 2 was that turning right rather than turning left would lead to a food outcome. This very same expectancy could be established in the same situation by changing o' from the water outcome of turning left to any other feasible differential outcome such as social contact with another rat, sexual commerce with a rat of the other sex, return to the home cage, etc.

Expectancies as Relations

Expectancies of the form $aa'Eo$ are not composed merely of a relation between a single act and its outcome; they involve a relation between *two* act-outcome sequences with a common outcome. Roughly speaking, an organism having such an expectancy may be said to relate the act-outcome sequence "a leads to o" to the sequence "a' leads to o" and to behave as if the former were more likely than the latter to be true. Earlier discussions of expectancies have not brought this relational character of the concept into focus. It is true that common speech permits the relation to go unexpressed, but, although one may say, "I expect to arrive in time if I take the train," it is clear that some alternative act is to be understood, such as, "rather than take a bus" or "rather than not taking the train" or the like. It should be noted that the alternative act is not necessarily the simple negative of the originally specified act.

If an organism can form such expectancies, it must also be able, at least implicitly, to form rules about temporal sequences of acts and outcomes. Such a supposition is natural to cognitive theory, the spirit of which is to understand the organism's behavior as determined in part by its knowledge of nature—at least, of that part of nature commonly referred to as its environment. It is also presumably meaningful to conceive of expectancies that do not refer to act-outcome sequences, such as the expectancies that it will rain tomorrow, that a friend will arrive on the five o'clock train, that a shock will be felt soon after a buzzer is heard, etc., but no attempt to deal with these will be made here.

Reversal of Preference

No doubt the reader will have felt that the requirement of experimentally reversing the preference relation in order to demonstrate expectancies is easier to demand than to accomplish. This feeling is justified; indeed, it can be argued that the fact that it is easier to reverse expectancies in order to demonstrate preferences than to reverse preferences in order to determine expectancies is a principal reason why there are many more experiments on the analysis of preferences than on the analysis of expectancies. Usually, expectancies are modified by giving the organism experience with a changed external situation, and preferences are varied by operations related to the biological state of the organism. Although the latter is often the harder task, increased knowledge, especially in physiological psychology, is already reducing the difficulty.

In many cases, of course, the reversal of preferences can be achieved by establishing contingent preferences between the outcomes under study and some pair of differential outcomes whose relative preference values for the organism are known and sufficiently definite. Nothing is more common in decision-making experiments than just this. It may be a matter of indifference to *S* whether a coin toss emerges as "heads" or "tails" or a drawn card is marked or blank or a key press produces a buzz or silence until he learns that money gains and losses are associated with these events. The criteria of expectancies given here require that *S* have preferences between the outcomes, but they do not require that these preferences be in some way intrinsic to the outcomes. No one (or, at any rate, no cognitively disposed psychologist) can doubt that Wolfe's (1936) chimpanzees formed various expectancies in which poker chips were the expected outcomes, but it is equally clear that the values of these tokens were formed by differential association with food, water, play, and other outcomes whose own values had a history that could only be surmised.

Strength of Expectancies

It is assumed that expectancies vary in their relative strength, but no formal treatment other than that of the third postulate stated above will be attempted here. It is obvious that to know that an expectancy exists is not to know its strength. For example, a child's expectancy that opening a blue box rather than a yellow one will gain him a toy may be weak, medium, or strong, and, in particular, to speak of the toy as the expectandum does not imply that the child sees the occurrence of this outcome as more likely than its nonoccurrence. Experiments could be arranged in which, although the probability of gaining a toy by opening the blue box was always greater than that of gaining the same toy by opening the yellow box, *both* probabilities were either higher or lower than .5, and the relative strength of the expectancies would presumably vary with these probabilities.

Expectancy Versus Drive Discrimination

If the rat of Illustration 2 is thought of as being switched between the two states, "hungry but not thirsty" and "thirsty but not hungry," and Outcomes *o* and *o′* are, respectively, food and water, then Table 1 is the design of certain "drive-discrimination" experiments, such as Leeper's (1935) main experiment and Kendler's (1946) test trials. In these instances (except for a few correction trials in Leeper's case), if the animal turned one way, it always found food and if it turned the other way, it always found water. This is a design for diagnosing expectancies, not discrimination; hence, the experiments are to be interpreted as demonstrating the existence of such expectancies as, say, that turning right rather than turning left will lead to food, and that turning left rather than turning right will lead to water.

Among other drive-discrimination experiments, those that used a correction method require special analysis and will not be classified here. These include Hull's (1933) original study of the problem and such recent experiments as those of Levine (1953), Bolles and Petrinovich (1954), and Deutsch (1959). A clear-cut discrimination design was, however, used by Jenkins and Hanratty (1949). When their rats turned right, they found food if they were less hungry and an empty box if they were more hungry ($11\frac{1}{2}$ versus $47\frac{1}{2}$ hours of food deprivation); when they turned left, they found food

TABLE 2

PARADIGM FOR DISCRIMINATION BETWEEN THE STATES oPo' AND $o'Po$

Alternative acts	Conditions			
	1 oPo'	2 $o'Po$	1' oPo'	2' $o'Po$
a	+ (m)	− (n)	− (n)	+ (m)
a'	− (n)	+ (m)	+ (m)	− (n)

Note.—m, n: differential outcomes with m preferred to n; oPo', $o'Po$: opposed preference states. Plus and minus signs indicate association between acts and preference states.

if they were more hungry and an empty box if they were less hungry. The outcomes of the acts were contingent upon the states of hunger that were used as discriminanda, thus fulfilling the fundamental criteria of a test of discrimination (Irwin, 1958). In general, tests of an organism's discrimination between two preference states such as oPo' and $o'Po$ require the paradigm of Table 2. The pattern of plus and minus signs in this table represents positive results of the test. The use of Outcomes m and n instead of o and o' in the cells of the table calls attention to the fact that tests of discrimination between two preference states can be made without employing outcomes relevant to those states. There must, of course, be a preference relation between m and n; in the use of the table, it is assumed that mPn. So, for example, to test for discrimination between the preferences for food over water and for water over food, one could let m be the turning off of an aversive bright light by pressing the correct one of two panels, and n be the null differential outcome of no change in illumination from pressing the incorrect panel. Had Levine (1953) and Bailey (1955) not used a correction method, this description would hold for their experiments.

COMPARISONS WITH OTHER FORMALIZATIONS

Tolman, Ritchie, and Kalish

Tolman et al. (1946, p. 15) wrote:

> DF. II. *If* x is deprived of food and x has been trained on path P and x is now put on path P and path P is now blocked and there are other paths which lead away from path P, one of which points directly to location L, *then* (x runs down the path which points directly to location $L \equiv x$ expects food at location L).

The present author's treatment of this situation would require several suppositions whose acceptability to Tolman et al. may be questionable. First, it would be supposed that "x is deprived of food" empirically implies that the animal desires food, that is, $fP\phi$.[2] It would further be supposed that the statement, "x runs down the path which points directly to location L," means that the animal chooses this act as against some other act, such as running down another path or not running at all. Finally, it would be supposed that, if food (f) could be made aversive to the animal, that is, if ϕPf, then the animal would perform the alternative act of not running or of choosing a different path. Given all of this, the results would fit Table 1, and the present author would accept the conclusion that the animal expected the act of running down the L-pointing path rather than committing the alternative act to lead to food, and also that the animal expected that committing the alternative act rather than running down the L-pointing path would lead to ϕ.

Two additional comments are necessary. The first has to do with the biconditional in the last clause of the Tolman et al. definition. Even if positive results in accordance with this definition permit the diagnosis of an expectancy, it is surely too strong to maintain that the converse is true, that is, that the existence of the expectancy necessarily leads to the appropriate choice behavior. Presumably Tolman et al. themselves would agree that the definition does not provide a statement of the sufficient conditions for the behavior. To do so would be to give a complete psychological theory, which the "definition" obviously does not do. The present author's defining criteria are subject to the same limitation. It is perfectly possible for an organism to have the required preference states and expectancies and yet, when introduced into the experimental situation, not make any choice at all, and the preferences and expectancies may be present when the organism is outside the experimental situation and therefore incapable of expressing them. If the behavior does not occur, there is no evidence, not negative evidence.

---

[2] See Footnote 1.

Second, the expectancy of Tolman et al. is defined under the condition that the animal has had a particular sort of training and has learned the expectancy during this training by a process of inference. But these are very restrictive conditions. On the one hand, it is more than possible that such expectancies might be formed by organisms that were incapable of drawing the inferences required by Tolman et al.; on the other, even granted such capability, the generality of such a conditional definition is far less than is desirable for a concept so fundamental to a cognitive view.

MacCorquodale and Meehl

In what is probably the most thorough attempt to formalize the notion of expectancy, MacCorquodale and Meehl (1953) set forth "preliminary suggestions" of a postulate system similar to Hull's but differing from it critically by substitution of "strength of expectancy," $(S_1R_1S_2)$, for Hull's "habit strength," ${}_SH_R$. The postulates provide for the growth of strength of expectancy by frequency of occurrence of the sequence $S_1 \rightarrow R_1 \rightarrow S_2$ when these elements are in close temporal contiguity, for weakening ("extinction") of the expectancy by occurrence of the sequence $S_1 \rightarrow R_1$ without its being followed by S_2, and for the development (by "inference") of an expectancy $(S_1R_1S^*)$ if the sequence S_2S^* occurs when $(S_1R_1S_2)$ already exists. The system is non-Hullian in that strengthening and weakening of expectancies by confirmation and disconfirmation take the place of growth of habits by reinforcement and extinction of responses by inhibition.

In comparing the MacCorquodale-Meehl expectancies with those described here, it must be kept in mind that the former are defined implicitly by their roles in a postulate system, whereas the latter are given explicit and quasi-independent definition by means of a canonical form of evidence. One consequence of this is that, while $(S_1R_1S_2)$ is defined in part by postulates that deal with its acquisition, $aa'Eo$ is defined without reference to this or any other "antecedent" condition. If this were the only major difference, it might be left for argument at the level of philosophy of science. There are, however, two other differences that appear to bear more directly upon the psychological nature of the processes under discussion: (*a*) the presence of the S_1 terms in the MacCorquodale-Meehl $(S_1R_1S_2)$ and the absence of corresponding terms from $aa'Eo$, and (*b*) the reference to an alternative act in $aa'Eo$ and the absence of this from $(S_1R_1S_2)$.

The first of these two differences, (*a*), is more apparent than real. It seems to do no violence to MacCorquodale and Meehl's $(S_1R_1S_2)$ to describe it as the expectancy that R_1 will lead to S_2, given that S_1 is originally present. Perhaps something is gained by the explicitness of this reference to the conditions under which the organism expects R_1 to result in S_2, but the present author's preference here, as also in connection with the problem of the relation between the definition of expectancies and their laws of acquisition, is to leave the matter for experiment to decide, rather than to put tight limitations on it at the level of postulate and definition. The present author supposes that the conditions under which such an expectancy as $aa'Eo$ will actually exist are among the very matters of fact that the psychologist should determine by experiment. Note that if the above translation of $(S_1R_1S_2)$ into words is accepted, then this is the same expectancy as $(S_nR_1S_2)$, where S_n is any arbitrarily chosen stimulus—the difference is merely that the expectancy R_1S_2 is contingent upon S_1 in the first case and upon S_n in the second. This does not seem to be incompatible in any way with MacCorquodale and Meehl's system, nor are these two expectancies incompatible with each other.

As far as (*b*) is concerned, the fact that $aa'Eo$ contains a reference to a pair of alternative, mutually exclusive acts, whereas $(S_1R_1S_2)$ does not, reflects at least a difference in the starting points of the schemes of which they form parts. The MacCorquodale-Meehl system shows its S–R paternity in that its variables are directed toward accounting for the quantitative properties of particular "responses." The final intermediary variable in the system is symbolized "$_SE_R$" and is spoken of as "the reaction potential . . . of a response R_1 in the presence of S_1 . . . [p. 58]." Neither explicit nor implicit reference to choice by the organism is found in the postulates except for the statement: "The oscillations of two different $_SE_R$'s are treated as independent, and the response which is momentarily 'ahead' is assumed to be emitted [p. 58]." On the other hand, the present paper sets out from a choice situation and makes the determination of one expectancy, $aa'Eo$, interdependent with the determination of another, $a'aEo'$.

Similarities and contrasts between the two conceptions can perhaps be seen best by asking whether there is an equivalence between the assertions "$aa'Eo$" and "$(S_1RS_2)/(S_1R'S_2) > 1$" (where R and R' are mutually exclusive responses), thus taking one of our expectancies to be analogous to the ratio of two of MacCorquodale and Meehl's expectancies. For the equivalence to hold, MacCorquodale and Meehl would have to maintain that "$(S_1RS_2)/(S_1R'S_2) > 1$" is materially equivalent to "$(\exists S_n)[(S_1R'S_n)/(S_1RS_n) > 1]$," in order to preserve the analogy to the material equivalence between "$aa'Eo$" and "$(\exists o')a'aEo'$." Such a claim seems inappropriate to their system and would at the very least require elimination of their assumption of independence among various $_SE_R$'s. This difference between the two sets of ideas is a direct and instructive consequence of the fact already mentioned, namely, that the one intends to study individual responses and the other intends to study phenomenon of choice.

One further point must be made. MacCorquodale and Meehl mention the difficulty experienced by Kendler (1946) in accounting for his own results without, to use their phrase, "smuggling in" the r_g–s_g mechanism. They attribute their own ease in handling the same results to the fact that their

> . . . *activation postulate makes reference to the expectandum.* . . . The expectancy frame retains the reference to the expectandum in the basic cognitive element (S_1RS_2) and thus a reference to this [expectandum] can be packed into the activation postulate in a way

that allows subsequent drive manipulations a ready control over the response strength [1953, pp. 59–60].

It seems that MacCorquodale and Meehl misplace their emphasis in these statements. It is important for "cognitive elements" of a theory of behavior to refer to outcomes of acts, but such reference is equally necessary in the *motivational* terms of the theory. MacCorquodale and Meehl provide exactly this in their activation postulate, the first sentence of which reads: "The reaction potential ${}_SE_R$ of a response R_1 in the presence of S_1 is a multiplicative function of the strength of the expectancy $(S_1R_1S_2)$ and the valence (retaining sign!) of the expectandum [p. 58]." Surely it is the reference to the outcome in "valence of the expectandum" rather than in the expectancy that is critical for the superiority of their system over S–R systems that employ, smuggled in or not, the r_g–s_g mechanism (cf. Deutsch, 1956). Spence (1956) maintains that s_g constitutes just such a reference to an outcome, and Osgood (1953, pp. 391–412) takes a similar attitude; but r_g–s_g resembles an expectancy rather than a preference. Such problems do not arise in the present system, since both the motivational and the cognitive processes that are involved in a choice refer to the same outcome. The chosen act is dependent upon a preference, oPo', and an expectancy, $aa'Eo$, that are interlocked through the identity of o, which is simultaneously the preferred outcome and the expectandum.

REFERENCES

BAILEY, C. J. The effectiveness of drives as cues. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 183–187.

BOLLES, R. C., & PETRINOVICH, L. A technique for obtaining rapid drive discrimination in the rat. *Journal of Comparative and Physiological Psychology*, 1954, **47**, 378–380.

DEUTSCH, J. A. The inadequacy of the Hullian derivations of reasoning and latent learning. *Psychological Review*, 1956, **63**, 389–399.

DEUTSCH, J. A. The Hull-Leeper drive discrimination situation—a control experiment. *Quarterly Journal of Experimental Psychology*, 1959, **11**, 155–163.

HULL, C. L. Differential habituation to internal stimuli in the albino rat. *Journal of Comparative Psychology*, 1933, **16**, 255–273.

IRWIN, F. W. An analysis of the concepts of *discrimination* and *preference*. *American Journal of Psychology*, 1958, **71**, 152–163.

IRWIN, F. W. On desire, aversion, and the affective zero. *Psychological Review*, 1961, **68**, 293–300.

JENKINS, J. J., & HANRATTY, J. A. Drive intensity discrimination in the albino rat. *Journal of Comparative and Physiological Psychology*, 1949, **42**, 228–232.

KENDLER, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *Journal of Experimental Psychology*, 1946, **36**, 212–220.

LEEPER, R. The role of motivation in learning: A study of the phenomenon of differential motivational control of the utilization of habits. *Journal of Genetic Psychology*, 1935, **46**, 3–40.

LEVINE, S. The role of irrelevant drive stimuli in learning. *Journal of Experimental Psychology*, 1953, **45**, 410–416.

MACCORQUODALE, K., & MEEHL, P. E. Preliminary suggestions as to a formalization of expectancy theory. *Psychological Review*, 1953, **60**, 55–63.

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford University Press, 1953.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

TOLMAN, E. C., RITCHIE, B. F., & KALISH, D. Studies in spatial learning: I. Orientation and the short-cut. *Journal of Experimental Psychology*, 1946, **36**, 13–24.

WOLFE, J. B. Effectiveness of token-rewards for chimpanzees. *Comparative Psychology Monographs*, 1936, **12**(5).

(Received May 6, 1965)

Psychological Review
1966, Vol. 73, No. 4, 335-351

NATURE OF THE EFFECT OF SET ON PERCEPTION[1]

RALPH NORMAN HABER

University of Rochester

To explain the effects of set on reports of perceptual experience 2 hypotheses are elaborated—set enhances the percept directly, or set facilitates responses and memorial organization of the perceptual experience that itself was uneffected by the set. Extensive research is reviewed that is relevant to each of these hypotheses, especially studies that attempt to differentiate them.

When a perceiver is given instruction to attend to only one of several attributes of a stimulus, he is able to report his perception of that attribute more accurately than he would be able to otherwise. Külpe, in 1904, first documented this effect experimentally, and, since then, a number of investigators have replicated the basic finding. The operations have typically involved either instructing the subject (*S*) as to which attribute he should attend or providing *S* with a set of alternatives, from which the stimulus has been drawn. However, the same process is implied by implicit instructions emanating from *S* himself, such as would occur when a hungry man is looking for food. The experimental demonstration requires that the set cannot be satisfied by peripheral mechanisms such as eye movements or scanning of the stimulus to locate the relevant attributes, since, otherwise, the phenomena would not be as interesting.

This paper is not a general review of the effects of set on perception, although such a review is badly needed to bring Gibson's (1941) discussion up to date. Rather, the topic here is an examination of the two basic hypotheses that have been used to interpret set effects. As these issues were not so clearly developed at the time of Gibson's review, little reference can be made to his paper.

At least two basic and dissimilar interpretations have been suggested to explain the effects of set on perception. The older one, favored by Külpe and most of the investigators following him, including the "New Look" theorists in perception, appeals to a perceptual enhancement or "tuning" hypothesis (Dember, 1960; Postman, 1963), whereby attending to a particular attribute of a stimulus results in a clearer and more vivid perception of that attribute—it stands out more. By the same token, the incidental attributes are not as clear and do not stand out. Thus, the perceptual tuning hypothesis places the locus of the effect of set in the perceptual system itself, occurring while the stimulus is being viewed.

The alternative hypothesis is that set has no effect on perception itself, but only on some aspects of the memory trace or on responses to that perceptual experience. Several varieties of this "response" hypothesis have been proposed: (*a*) The set facilitates relevant responses by *S*, increasing the probability that *S* can identify the stimulus in his report; (*b*) the set causes *S* to report the emphasized attributes first, before memory of the stimulus fades, thus allowing those attributes to be reported more accurately; (*c*) the set modifies the organization of the memory trace, so that the important attributes are remembered more accurately.

---

[1] Partial support for this research was provided by a grant from the United States Public Health Service (MH 03244) to the author at Yale University, where this work had been completed. I am most appreciative to Thomas Natsoulas and Maurice Hershenson for their many valuable comments on the manuscript.

The two basic hypotheses have wider application in perception and memory than just to problems of set. Behind them lies specification of one of the most difficult distinctions in any analysis of perceptual behavior. The tendency has been to assume that just because *S* was reporting immediately what he had perceived he was therefore performing a "perceptual" task. There is a crucial difference between reporting one's experience and reporting the attributes of stimuli that one remembers seeing. One task is perceptual, while the other deals with a memorial process. These must be conceptually differentiated, even if they may function similarly, and even if they proceed originally from the same sensory base. There are innumerable experiments that fail to make this distinction, and much of the controversy among perceptual theories results directly from this confusion. The two hypotheses under discussion in this paper reflect the distinction being proposed. To provide evidence relating to either or both of these requires that operational definitions be made for measurement of perceptual experience directly, or that deductions from these hypotheses specify operations in terms of memory alone that will still permit inferences concerning perceptual experience. As the present discussion will amply illustrate, neither of these conditions is well met in practice, though abundant data are available regarding both of these hypotheses.

Needless to say, the earliest experiments (e.g., Külpe, 1904; Rubin, 1913; Wilcocks, 1925; Yokahama, reported in Boring, 1924) were not designed to separate these two basis hypotheses, even though the hypotheses were discussed as alternative explanations for the various results.

Külpe's (1904) stimuli were groups of letters varying in color and location. Prior to most presentations, *S* was given a set to note only one of the several attributes of the stimulus: the number, position, or color of the letters, or the overall figure formed by the letters. For the remaining stimuli, no specific set was given. After the presentation, *S* was instructed to report the attributes as best he could, with no order specified. If he left any out, the experimenter (*E*) questioned him on those. Külpe found that *S* was more accurate in reporting the attribute for which he had paid attention prior to the presentation and was less accurate in reporting the others. Külpe suggested that the attention instructions facilitated the clarity of the percept.

Wilcocks (1925), responding to a criticism of Külpe by Rubin (1913), extended Külpe's method to include a memory variable. Wilcocks argued that if the reports of attributes after the presentation were in an uncontrolled order, it is possible that *S* might have forgotten some of them by the time he was questioned. Therefore, Wilcocks randomly arranged the order of reporting the attributes, though, in other respects, his experiment was similar to Külpe's. He found in general that accuracy was poorer the later an attribute was reported. Even so, he found that when order of report was controlled, the report of the attended-to attribute was more accurate than that of the unattended-to ones. Thus, while a memory variable was shown to be operative, it seemed to be independent of the enhancement effect. Therefore, Wilcocks concluded that Külpe's basic finding was replicated and could not be reduced to a memory process. Likewise, he accepted Külpe's interpretation of this process as perceptual enhancement.

Chapman (1932) used similar figures, where *S* could either count the letters, locate their positions, or name them. Instructions to attend to only one of these attributes were given in some sessions 4 seconds prior to the stimulus exposure, and in the remaining sessions 2 seconds following the exposure, in both cases by means of an illuminated display. Chapman's results showed that accuracy of report was higher when the attention instruction was given before the stimulus exposure, as compared to afterwards. Comments made by *S*s indicated that the attended-to attribute was perceived more clearly during the actual exposure of the stimulus. They felt they did worse in the aftercondition because they had to keep too many things in mind. However, the comments also suggested that *S*s' memory of the stimulus was affected by the type of set given. In the prior condition, they indicated that the memory for the irrelevant attributes faded and became indistinct. Thus, Chapman concluded that attention instructions given prior to the exposure both enhanced the perception of the stimulus itself during its actual exposure and, further, started in motion a differential forgetting process, where the relevant attribute was maintained in memory more accurately and for a longer time.

The evidence regarding the effect of set reviewed to this point supports a perceptual tuning hypothesis, primarily on the basis of *S*s' postexperimental introspective comments. They report that the set-for attributes are clearer, stand out more, fade less rapidly, and the like. Support for at least one of the response hypotheses is also found, based on before-after differences, order of report effects, as well as introspective comments regarding fading memory. In light of the importance of an *S*'s report of his perceptual experience, the comments regarding experience in these experiments must be given weight, even though they were not collected in any systematic or complete fashion nor subjected to any content analysis, and even though *S* was given no pretraining on the categories to use in his report. (See Postman, 1963, for a discussion of these criteria to be used for the evaluation of introspective reports in perceptual research.)

While an *S* in this type of experiment might be confident that he noticed changes in the perceptual characteristics of the stimulus as his set changed, no converging operations (Garner, Hake, & Eriksen, 1956) sufficient to demonstrate the validity of those reports were employed. Introspectively, it might be very difficult for an *S* to differentiate a rapidly fading memory, for example, from an image that was perceptually less clear. Unless experiments are designed to guarantee that *S* can make this distinction introspectively (and the ones just reviewed clearly do not offer such a guarantee), they cannot provide unequivocal evidence regarding either of these two hypotheses. Therefore, this evidence can be considered only suggestive support for the perceptual tuning hypotheses, as well as providing some evidence for at least one of the versions of the response hypothesis.

The New Look in perception, beginning after World War II, accepted this evidence, however, as offering strong support for the perceptual tuning hypothesis (Bruner, 1957a, 1957b). A number of experiments were then reported that seemed to provide further support for this hypothesis, including ones by Neisser (1954), Postman and Bruner (1949), Bruner and Postman (1949), Krulee, Podell, and Ronco (1954), Green and Anderson (1956), Ross, Yarczower, and Williams (1956), Hoisington and Spencer (1958), to name only a few. For example, Postman and Bruner (1949) showed that if *S* was given a multiple set of alternatives (e.g., one of two stimulus words will be either a food or a color word—report only that one), he had higher recognition thresholds than if a single set was given (e.g., one of two stimulus words will be a food word—report only that one). Thus, they argued that since the recognition thresholds were manipulated by the set, the set had actually changed the perceptual characteristics of the stimulus.

Neisser (1954) presented *S*s a list of words to study, indicating that some of them would later be shown tachistoscopically. He did in fact show some of the original words, but, in addition, exposed some that were homonyms of other words on the list. He found no facilitation of the homonyms, even though the original words were reported more rapidly. Since a word and its homonym require the same response, but only the original word was facilitated, he concluded that the presentation of the actual word prior to exposure must have made the word itself stand out perceptually.

Bruner and Postman (1949) gave *S*s the set to expect to see normal playing cards, but in fact showed them cards in which the colors and suits did not agree (e.g., black hearts). While many *S*s apparently ignored the incongruity and reported the cards as they would normally appear (e.g., red hearts or black clubs or spades), a few *S*s reported seeing "compromise" perceptions (e.g., brown or purple hearts). These latter responses have been interpreted as supporting a perceptual hypothesis, since no response facilitation could be expected for a category such as purple hearts that *S* had never seen before.

Bruner (1957b) has offered an intriguing suggestion regarding the relationship between the two hypotheses under discussion. He proposes that perceptual experience (the concern of the first hypothesis) is the result or end product of categorization (the concern of the second hypothesis). Thus, an object is experienced or perceived *only after* it has been properly classified—that is, responded to. Prior to its identification, the perceptual experience of a stimulus is blurred and indistinct. In the present context, then, if the perceiver has a set which facilitates classification of the stimulus, it also then facilitates the clarity of its percept.

Bruner and Potter (1964) present some evidence relevant to one aspect of this notion. They show *S* a common, everyday object extremely out of focus and then gradually increase the focus until it is perfectly clear. During the period of increasing focus, *S* is required to give a running commentary of what he thinks the picture contains, from which the point of correct identification can be determined. They found that if the focus increases very slowly, so that *S* has extensive viewing during which he cannot correctly classify the stimulus, he requires a greater amount of focus before correct identification occurs. From this the authors argued that the extent to which *S* misclassifies the picture during the very fuzzy presentations prolongs the time before he can see it clearly. Presumably, during this time, *S* is testing his classifications against the fuzzy image, and he has to reject the wrong classifications he made before he can concentrate on testing correct ones. Only after he tries and checks the correct one does he come to see the picture clearly.

Obviously, this kind of experiment, interesting as it is, cannot provide much evidence for as complex a position as Bruner is espousing here. He has to make too many additional assumptions to have a clear-cut analysis. A somewhat more relevant, though still indirect, test of this notion was reported by Hershenson and Haber (1965). These authors had previously found that the clarity of letters of words increased markedly if they were flashed several times, even when on the first flash *S* could see none or few of the letters (Haber, 1965a; Haber & Hershenson, 1965). However, the rate of increase in clarity for Turkish words was slightly faster than that for English words. It was assumed that *S* would be less likely to misclassify what he knew was a Turkish word from only partial information, since he knew it would be a totally unfamiliar word. Therefore, this result was interpreted to imply that when *S* made incorrect anticipatory classificatory responses for the English words, this slowed down the development of their clarity.

Even given these kinds of data, Bruner's notion faces serious theoretical problems on other counts, as, for example, how one can see something that is novel when he cannot classify it correctly. Therefore, until more than this meager and indirect evidence is available, little else can be said for this version of a linkage between the two processes.

An independent line of research, originally led by Ames (e.g., 1951), has provided the basis of the Transactionalist theory of perception. In this context, many demonstrations were created to show how an *S*'s assumptions based on past experience or knowledge directly affect his perceptual experience. In one of the most striking of these, a physically rotating trapezoidal-shaped window appears to oscillate back and forth, and to be rectangular in shape. This is a very powerful perceptual experience, experienced by nearly every viewer, so that it is difficult to explain in terms of any kind of response process. Haber (1965b) has recently shown that if *S*s are given extensive experience with the apparatus and knowledge of the theory behind the illusion, their perceptual experience changes so the majority of *S*s report seeing the nonillusory rotation of the trapezoidal figure. Numerous other examples of these demonstrations can also be used to support the effects of *S*'s assumptions (set) on his perceptual experience directly, in ways suggesting that response processes are inoperative.

Shortly after the beginnings of the New Look, strong interest was revived in alternative interpretations involving response and memory processes. Many of the New Look experiments were scrutinized for evidence of these processes or for faults in design. Accompanying this methodological concern was the new development in theorizing about methods of research in perception. The most important single paper in this regard is by Garner et al. (1956). While not taking sides with respect to these two hypotheses, Garner et al. point out quite persuasively the inadequacy of nearly all previous research to decide among the interpretations of the effects of set, as well as motivation and implicit instructions, on perception. They suggest several nonperceptual interpretations of these patterns of results and present the concept of converging operations as an experimental design to provide tests between the two hypotheses. They are particularly concerned with the possibility that the set limits the number of alternatives from which *S* must make his response, thus increasing the probability that the correct response will be made, independent of any perceptual enhancement. Eriksen (1958, 1960, 1962) has continued this position, arguing that nearly all research in this area has failed to differentiate the two hypotheses satisfactorily, and that since from his point of view the response interpretations are more parsimonious, there is no evidence to support a perceptual interpretation. His position has received added support from a number of recent studies.

Lawrence and Coles (1954) argued, along with Chapman (1932), that a perceptual tuning hypothesis has to predict that a set given before the stimulus, as compared to one given afterwards, will provide greater facilitation of report. They go further and suggest a dual hypothesis: Perhaps perceptual enhancement does occur if a prior set is given, but, when given afterwards, facilitation could be due to some selectivity in the memory process. If this were true, then an experiment like Chapman's would not distinguish between the purely perceptual and the dual hypothesis, since they might both be correct.

Lawrence and Coles (1954) used pictures of single familiar objects, with the set administered by projecting on the screen, either immediately before or after the stimulus exposure, the names of four alternatives for the stimulus. The alternatives were very similar to one another for some trials, while on others they were quite discrete. One group of *S*s had no alternatives at all; this served as a control on the use of the alternatives.

The results showed that giving alternatives of either type increased the accuracy of recognition of the stimuli, as compared with the no-alternative control group. Hence, the effect of set was demonstrated. However, alternatives given after the exposure were just as useful as those given before (which stands in contrast to Chapman's before-after difference).

This failure to replicate Chapman's before-after difference was not considered crucial, since the equality may have been due to perceptual enhancement from the before set and memory selection from the after set—the two processes just balancing each other. However, they also failed to find any interaction between the before-after sets and the type of alternatives (similar or discrete). They had argued that a perceptual hypothesis would predict no difference between the two types of alternatives when they are given before the presentation, since the correct alternative would stand out whether other alternatives were similar or not. On the other hand, since the after set could not affect enhancement, it must affect memory, which should show a differential effect of the two types of alternatives—discrete ones being more effective in facilitating correct recognition than similar ones. Failure to find evidence in support of this prediction (discrete alternatives consistently facilitated recognition more than similar ones for both before and after sets) led Lawrence and Coles (1954) to reject any form of the perceptual tuning hypothesis, and to argue that the effect of set was via some selective facilitation of memory processes.

By implication, Lawrence and Coles provided a possible explanation for Chapman's before-after difference—the one they failed to replicate. Under the assumption that the entire effect of set is caused by memory and response processes, the earlier the set is given, the greater the opportunity it will have to provide facilitation of these processes. Thus, even if there is no perceptual tuning, administering the set prior to the exposures gives *S* just that much more time to organize his memory. Sets given after the stimulus presentation have to act on an already fading memory trace and are at a disadvantage compared to a set given before the trace has weakened. As the delay between the exposure and the set increases, the superiority of the prior alternatives should increase. Since Chapman had a longer delay than Lawrence and Coles, this slight time difference may have accounted for the different outcome in the two experiments.

An interrelated group of studies by Long, Reid, and Henneman (1960), Long, Henneman, and Garvey (1960), and Reid, Henneman, and Long (1960) has used a model of set as a response limiter, similar to Garner et al.'s (1956) notion, whereby set "increases perceptual accuracy by increasing the probabilities of certain responses and decreases the probabilities of others in the various response classes [Long, Reid, & Henneman, 1960, p. 554]." Most of their experiments used single letters as stimuli, degraded in clarity, with set manipulated by giving *S* a group of alternative letters, of which the stimulus letter would be one. The alternatives were given either before or after the exposure. They found that decreasing the number of alternatives increased the accuracy, regardless of whether the alternatives were given before or after the exposure. From this they concluded, as did Lawrence and Coles (1954), that set was acting on the selection of responses, and not on perception itself. Greater accuracy was also found if the alternatives were drawn from an exclusive subset rather than from the entire alphabet. Further, they showed that set effects could be produced by restriction of the categories (rather than alternatives) into which the stimulus might fall. This is somewhat similar to the Postman and Bruner (1949) finding, but was found to be true regardless of whether the restriction was given before or after the stimulus. This greatly weakens Postman and Bruner's contention that restriction of category affects the perception of the stimulus itself. All of these findings are consistent with a response-limiting interpretation of set.

However, Long, Reid, and Henneman (1960) also found and replicated several before-after differences. One was found when only two alternatives were provided, so that being told the stimulus was one of two particular letters before the flash helped *S* report the letter more accurately than if such a set was given 1 or 2 seconds after the flash. Another before-after difference occurred in an auditory version of their experiments, when four alternatives were presented either before or after the stimulus letter was spoken in noise. Finally, when two degraded letters were presented and *S* was asked which of the two was a specific alternative, they found that specifying the search letter before the presentation facilitated correct discrimination, as compared to a later specification.

Each of these findings is inconsistent with the response interpretations of set, and it is not clear how they might be explained other than by some form of perceptual enhancement. The Lawrence and Coles' explanation—that the earlier the set is given, the greater its opportunity to facilitate memory processes—will not work, since it should follow that the greater the number of alternatives, the greater the effect of an earlier set. "Increasing the alternatives implies that the number of discriminating characteristics between them become more restrictive, and thus are more likely to be lost in the trace [Lawrence & Coles, 1954, p. 213]." However, these findings, inconsistent with a response interpretation, occurred primarily with a few alternatives, and no before-after differences were found with larger number of alternatives.

In a different line of research, Taub (1965), following the work of Teichner, Reilly, and Sadler (1961) and Teichner and Sadler (1962), presented an array of letters to *S*, asking him to report all of them. The number of letters in the array, the exposure duration, and the relative value of each of the letters (set) were varied. Value was manipulated by assigning points for correct reports of letters, with differential value attached to letters from the first and second half of the alphabet. Taub found that value affected accuracy (a set effect), but primarily through a reduction in accuracy of the low-valued letters, rather than an increase in the high-valued ones. This occurred in spite of the finding that the first letter reported from an array was most likely to be one of high value.

Taub's (1965) experiment is most relevant to the issues of this discussion because of the similarity of a prediction he makes to the Lawrence and Coles' (1954) similar-discrete difference presented earlier:

> . . . if a major selective process in perception was involved, then the size of the difference between high- and low-value letters should vary inversely with length of exposure. That is, a perceptual set, which would make the higher value categories easier and quicker to identify (Bruner, 1957), would occur at those conditions where there is little time to detect and discriminate between the categories, but have little effect at longer exposures where all letters would be equally discriminable. Predictions from a selective recall point of view would be the opposite. That is, as exposure time increases and the rate of memory loading decreases, *S*s would have more time to selectively code the categories which are detected. Thus, a selective memory process explanation would predict increasing differences in response to value as exposure time increases [p. 142].

Since Taub found unequivocal evidence that increasing exposure time increased the differences in accuracy between high- and low-value words, he rejects a perceptual tuning explanation in favor of some variety of the memory hypothesis—especially a selective-recall one. However, Taub used relatively long exposure durations (500–2,500 milliseconds), so that visual searching clearly occurred for the longer presentations. Therefore, his data are equivocal with respect to this issue, since *S* has opportunity to seek out the high-value items if given sufficient time, and this would imply nothing about the effects on perception at all. Thus, it appears that, while the prediction is relevant, it requires a presentation time considerably shorter than Taub used to test it adequately.

A different type of experiment was reported by Lawrence and LaBerge (1956) in which they were concerned with the response hypothesis of a fading memory. The stimuli were pairs of cards from the Wisconsin Card Sorting Task, with an attention-instruction set given only before the exposures, by indicating the relative value for each attribute correctly reported. For one group of *S*s, all three attributes (number, color, shape) received equal value; for another, one attribute was always emphasized by offering 100 times as much money as each of the other two attributes. All three attributes had to be reported, but *S* determined the order. A third group, with equal set, had the order in which the three attributes were to be reported controlled and varied from trial to trial.

When the emphasis and equal conditions were compared, they found that the effect of set appeared to increase accuracy for the attended-to attribute and decrease it for the other two. They also found that the decrement from first to third report under the equal condition was about the same magnitude as the difference between the critical and incidental attributes in the emphasis condition. While they did not measure the order of report when unequal attention instructions had been given, Lawrence and LaBerge (1956) argued that the set effect could have been entirely determined by the order in which *S* reported the attributes. That is, if, whenever one attribute was critical, *S* always reported that one first, then the accuracy of the critical attribute would be higher because it had been reported first.

Thus, Lawrence and LaBerge suggest that no assumption need be made about reorganization of memory or changes in response probabilities, but only that all memory for the stimulus is slowly fading. Whatever is reported first will be more accurate than later-reported items. This hypothesis is very similar to one Wilcocks (1925) tested but rejected. Lawrence and LaBerge cannot test this in their data since they did not measure the preferred order of report when the attention instructions made one attribute critical.

This test was provided in experiments by Harris and Haber (1963) and Haber (1964a). The same stimulus materials and procedures for inducing sets were used, but each *S* was tested under all set conditions, including a forced order of report specified by *E* on half of the trials, and a free order on the other half. *S* could report the stimulus in whatever order he chose (though he had to indicate the one he used). Under forced order of report, when just the first attribute reported was analyzed, if that first report was of a critical attribute, it was more accurately reported than if it was of an incidental one. The same was true for the second and third reports. Thus, Lawrence and LaBerge (1956) were incorrect when they argued that the set effect could be explained in terms of order of report. Holding order of report constant, the set effect was still obtained. Harris and Haber (1963) did find that some *S*s reported the critical attributes first if they were allowed to choose their own order of report. But even granting this, it did not account for the set effect. Thus, this simpler hypothesis of Lawrence and LaBerge was shown to be inadequate.

While the findings of Harris and Haber (1963) and Haber (1964a) indicated that fading memory coupled with a varying order of report could not explain the set effect, they did discover what may be a more explicit determiner of set—namely, the encoding processes by which *S* translates the percept of the stimulus into memory. Some pilot work had indicated that nearly all *S*s spontaneously encoded this type of stimulus into words and silently rehearsed the message repeatedly until all of the attributes had been reported. Further, with these stimuli, it was found that *S*s used one of two dissimilar strategies to accomplish this encoding. Some *S*s encoded it by separating the stimulus into objects on the left and right (e.g., one red triangle, three blue stars). The remaining *S*s encoded it by separating the stimulus into its three dimensions (e.g., red blue, triangle star, one three). The order of the dimensions in this latter Dimensions code could of course be varied without disturbing the strategy, while the former Objects coding strategy has its order of encoding fixed by the rules of English syntax.

To investigate the effects of differential coding strategy on perception, Harris and Haber (1963) and Haber (1964a) trained *S*s to use one of the two coding strategies. In both of these experiments, an effect of set and of order of report was found. (Emphasized dimensions were reported more accurately than unemphasized dimensions, holding order of report constant, and the dimensions reported first were more accurate than those reported later.) However, both of these effects strongly interacted with coding strategy. So strongly, in fact, that nearly all of these effects were accounted for by the performance of those *S*s using the Dimensions code. Thus, those *S*s who used the Objects code, based on English syntax, were not significantly more accurate on the emphasized, as compared to the unemphasized, dimensions. Further, regarding the order of report effect, as shown in the replication study (Haber, 1964a), the difference between the early and later reported dimensions was smaller for those *S*s using the Objects code than for those using the Dimensions code. Finally, *S*s using the Objects code had a greater overall accuracy of report, regardless of condition.

It seems clear therefore that the set and order of report effect can in large part be explained by the operations of the coding strategies themselves. These results suggest that the set effect stems as much from a change in the initial organization of the memory of the stimulus produced by the strategy by which *S* encodes the stimulus, rather than from a facilitation of contact with an already organized memory or through the effects of a fading memory or through the changes in the probability of the correct response being available.

However, the results on coding strategy do not indicate the reasons for the differences between these two coding strategies. To assess these, Haber (1964b) conducted a third study, using the same stimuli, strategy-training procedures, and set manipulation, but requiring *S* to verbalize his encoding and rehearsal of the encoded stimulus so it could be tape-recorded. Each *S* was instructed to begin talking immediately after the flash with his verbal encoding of the stimulus, and to repeat this encoding over and over for 20 seconds.

Haber found that *S*s using the Objects code could begin encoding considerably faster and complete encoding more rapidly than *S*s using the Dimensions code. Thus, the two strategies differed sharply in the speed with which they could translate the stimulus into memory. If there is some afterimage or short-term memory of the stimulus, lasting perhaps several hundred milliseconds beyond its offset (Averbach & Coriell, 1961; Mackworth, 1963; Sperling, 1960), then the strategy that can translate this image into memory most rapidly is more likely to have that encoding based on a good image that will allow it to be more accurate. In this way, one can account for the difference in overall accuracy between the strategies. Similarly, if one strategy encodes slower than the other, then presumably the items it encoded last would be poorer in accuracy, which would suggest that the order of report effect could also be accounted for by the differences in the speed of encoding. Finally, it was found that the Dimensions coders nearly always encoded the critical dimension first, while the Objects coders always followed the same order—that of English syntax. Thus, the set variable can also be explained by the differential speed of encoding. If the attributes encoded first are always the critical ones, leaving the others to be encoded later from a rapidly fading image of the stimulus, it is not surprising that the critical attribute is reported more accurately. This does not depend upon order of *report*, as Lawrence and LaBerge thought, but on order of *encoding*.

The two strategies also differed in

the frequency of errors made after the initial encoding, with transpositions, intrusions, and omissions occurring significantly more frequently during the rehearsal process in the Dimensions code. This could further account for the order of report effect—the items reported later would be based upon a rehearsed encoding that had undergone some erroneous change. This would be true whether *S* was reporting the stimulus attributes in an order he chose or in one forced upon him by the experiment.

This experiment suggests, then, that the coding strategies differ in two important ways: the speed of encoding. and the resistance to errors during rehearsal. It further suggests that these two differences could account for both the effect of set and the effect of the order of report in the experiments by Lawrence and LaBerge (1956), Harris and Haber (1963), and Haber (1964a), as well as by implication some of the earlier experiments discussed above. For example, the study by Neisser (1954) can now be explained in terms of encoding, without having to make any assumptions about the nature of the percept. This alternative explanation suggests that while the response given to the homonyms NO and KNOW is identical, the encoding of those two words into memory would be quite different. One reason for the difference could be that encoding is faster for words one expects to see. Thus, encoding of the unexpected KNOW begins later, takes longer, and therefore uses a progressively more degraded trace. While there has been no direct test of this explanation, it seems reasonable that the more prepared *S* is for what he will see, the more likely it is that he will have an appropriate strategy and category available in which to encode the stimulus.

It should be noted that variation in encoding is not the only nonperceptual interpretation of the Neisser (1954) experiment and of experiments similar to it. Since Neisser's *S*s are in a guessing situation, where they must guess the word from presumably only fragments of the letters given perceptually (at least on the first few flashes), they will have more success guessing those words they were given a set to see (e.g., NO will be easier to guess than its homonym KNOW, which they had not experienced in the preliminary training). This interpretation has been used by Eriksen and Browne (1956) in a related type of study. It says nothing about increased clarity of the stimulus caused by the prior set, but only that such a set may facilitate guessing of the correct word when only a few pieces of it are seen perceptually.

The encoding analysis advanced above is a more specific hypothesis about the outcome of the Neisser experiment than the more general guessing one. It attempts to specify the translation of the visual image, as it is generated on the retina by the stimulus, into some kind of memorial trace or persistence after the stimulus terminates. While the visual image may persist for a brief time after the stimulus offset, in the form of a short-term memory, it is crucial that perceptual theory take into account the processes that occur after that short-term memory has faded. Since it is obvious that memory for stimuli persists long after the stimulus (and any attendant short-term memory and afterimages) has terminated, no perceptual theory can be complete without cognizance of this persistance of memory.

Specifically, the encoding explanation being discussed here suggests that most visual stimuli are remembered by being encoded into previously learned linguistic units, usually words. This encoding takes place while the stimulus (or its brief short-term memory) is still present, and whatever has not been encoded after the stimulus has faded is entirely lost and not available in permanent memory. This is not to deny that a few *S*s may use some nonlinguistic code, especially for nonlinguistic stimuli. However, certainly with respect to the kinds of materials of greatest interest to psychologists, encoding into words is the most probable basic strategy.

Variation in the adequacy of encoding can presumably depend upon a number of variables. The experiments reported above on encoding of Wisconsin Card Sorting Task concept cards suggest that when a number of highly familiar dimensions of a stimulus must be encoded quickly, accuracy of encoding depends upon whether *S* has at his disposal a strategy that lends itself to rapid coding, preferably without having both to learn the code as well as perform the encoding operations. Thus, using a highly overlearned syntax for the order of encoding is preferable to one for which *S* must think about what should be encoded next.

In many perceptual experiments, the problem is not one of speed of encoding familiar dimensions, but of encoding degraded and often unpredictable stimuli. For example, a typical trial of a word-recognition experiment presents *S* with a fragmentary flash which first has to be maintained in memory and then that memory compared with possible alternatives of what the original stimulus must have been. If *S* has no adequate way of maintaining the fragments in memory, he will not be able to test any hypotheses about the stimuli once it terminates.

Thus, in most perceptual situations, two sources of variation are present—the adequacy of encoding a visual image, often imperfect, into memory, and the success of guessing what stimulus gave rise to those remembered fragments. To take an extreme though not too uncommon instance, stimuli for which *S* has no code for translation into memory are notoriously difficult to remember, even though *S* may perform perfectly on simultaneous discriminations with such stimuli. That is, he can make all of the appropriate responses to the stimulus while it is present, but cannot reconstruct it after it terminates. Most people experience this difficulty in describing colors, tastes, odors, and feelings from recollection. It is usually suggested that they do not have the appropriate words or other codes with which to maintain them in memory. These examples represent instances of excellent matching between the stimulus and prior experience, but very poor encoding of the stimulus into memory. Thus, short-duration presentations or delayed responses would show poor performance. One exception to this is found for those few people with eidetic imagery (see Haber & Haber, 1964) who are capable of maintaining a nearly perfect visual image of the stimulus for many minutes. These eidetic perceivers seem to do little if any encoding of the stimulus while the image persists, so that once it fades their memory of the stimulus is little better than that of more typical individuals.

Evidence relating to this specific distinction between encoding the percept and then matching the resultant memory to prior knowledge is very sparse, though it is beginning to accumulate. For example, McKinney (in press), by presenting single letters continuously and asking *S* to report changes or fragmentation in the percept of the letter, found that if the letter was embedded in a series of other letters which *S* was labeling as letters, much less change was reported than if it was embedded in a series of geometric designs for which no labels were suggested. Thus, the stability of the percept is increased when the object can be labeled (an operation probably equivalent to encoding).

Hintzman (1965), following work by Conrad (1964), presented lists of 8 symbols, selected so that matched pairs of them would be auditorily confusable (e.g., 2 with Q, 3 with T, etc.). Analyses of errors showed that auditory confusions were far greater than chance, suggesting that *S*s were maintaining these symbols in memory by some type of auditory encoding and rehearsal. This was explicitly found when, for the half of the trials on which white noise was used to prevent *S* from hearing his own thoughts, *S* produced overt rehearsal. However, Hintzman also reports that some *S*s did not use rote aural encoding, but rather attempted to classify the string of symbols (e.g., to make up a mathematical formula). Such *S*s made few aural confusions. Thus, the nature of the coding adopted by *S* determined the pattern of errors made. This, of course, was a situation where the symbols from which each presentation was drawn were well known in advance. Hence the matching of an encoded memory to prior experience was presumably perfect, with all of the variance in the encoding of that memory.

Gruber, Kulkin, and Schwartz (1965) presented a paired-associate learning task in which half of the *S*s were instructed to attempt to form a visual image between the two members of each pair. These *S*s had higher recall scores for all exposure durations, ranging from 1–24 seconds. Here, again, it would seem that aids for encoding of perceptually presented familiar material increase the accuracy of its encoded representation in memory.

The investigation and manipulation of coding variables in general in the study of perceptual and cognitive processes is becoming more extensive and is playing a far greater role in theory than 5–10 years ago. Much of the impetus for this has been Miller's (1956) paper on coding effects in information processing, although scattered work was reported before (e.g., Fitts, 1954; Pollack, 1952). More recently, however, a number of important experimental and theoretical papers have applied encoding processes to a wide range of perceptual data and problems (e.g., Broadbent, 1958; Conrad, 1964; Glanzer & Clark, 1962, 1963; Mackworth, 1963; Miller, Galanter, & Pribrim, 1960; Sperling, 1963).

The three studies on coding strategies by Haber have shown the importance of coding processes in the investigation of the effects of set. They suggest the possibility that many of the effects of set on reports of perception can be explained by differential encoding processes. Many studies have reported that the emphasized attribute of a stimulus can be reported more accurately than other attributes, irrespective of when the instruction for emphasis is given. However, in the Haber studies, this finding is dependent upon the nature of the encoding strategy—only one type of strategy will produce this emphasis-set effect. Similarly, only one type of strategy leads to an order of report effect. Thus, whether attention instructions (and restriction of alternatives) will lead to a set effect may depend upon whether *S*'s encoding strategy is susceptible to such a manipulation.

An alternative possibility is that on the trials when the set is given to *S*, he changes his strategy so as to maximize the payoff from the set. In the Haber experiments, only a few *S*s varied their strategies from condition to condition—too few to provide any kind of analysis of their performance. But this is a very likely possibility, particularly if the various set conditions make quite different demands upon *S*s' reports. Since the studies by Haber are the only ones that have shown this higher-order dependency of set on encoding, further research will be needed before these possibilities can be analyzed successfully.

CONCLUSIONS

Evidence on the effects of set has been reviewed with respect to two basically different hypotheses—set enhances the percept of the stimulus while *S* is actually viewing it; set facilitates report of the stimulus without affecting its percept.

At least three varieties of nonperceptual hypotheses were discussed: (*a*) response-limiting or response-probability changes, (*b*) order of report changes coupled with a fading memory, and (*c*) reorganization of the memory process itself. The second alternative does not seem to be a necessary condition for the production of set effects, since even when order of report is controlled the effects of set are still found without loss in magnitude. The supporting evidence for the response-probability interpretation is extensive, in that the magnitude of the set effect varies with the manipulation of the probabilities of responses or limitations on responses. The third alternative, reorganization of the memory process mediated by *S*'s coding strategy, is strongly supported in the results of three studies by Harris and Haber (1963) and Haber (1964a, 1964b). Further, Haber's interpretations of those results suggest that the response-probability explanation may be reduced to memory reorganization (encoding), so that only one nonperceptual hypothesis may be needed.

While this review then suggests a narrowing of the number of nonperceptual alternatives, it provides less resolution to which of the two basic hypotheses is correct. The problem of analysis is still very complex—many of the experiments discussed provided clear-cut evidence for response or memory explanations, but without simultaneously demonstrating a lack of a perceptual effect. That is, finding evidence in favor of one hypothesis does not disprove the other, since they are not incompatible, only alternative. There is no reason to doubt that both may be correct and occurring together.

It is likely that a set effect created by differential value (e.g., Harris & Haber, 1963; Külpe, 1904; Lawrence & LaBerge, 1956) is mediated primarily or perhaps totally by characteristics of *S*'s coding strategy. However, even with this, the Harris and Haber (1963) and Haber (1964a, 1964b) experiments do not rule out perceptual changes, since they had no converging operations sufficient to separate percept from coding effects. Their argument took the form that the set effect was present only in *S*s using one type of strategy and not in *S*s using another—therefore the set effect must have been due to properties of the strategy. Further, when the strategies themselves were examined, temporal and interference processes were found that were sufficient to account for the differential set effects between the two strategies, though again, this does not rule out perceptual effects completely.

So what conclusions does this evidence permit? Most *S*s feel an increase in perceptual clarity occurs under appropriate set conditions, though their reports give no guarantee that they can really distinguish felt clarity from a better memory. On the other hand, there is no reason to doubt their reports either. Just because the evidence supporting response processes is clear does not imply that perceptual enhancement does not occur. Further, Bruner and Postman's (1949) finding of compromise perceptions and the strength of the trapezoidal illusion, to pick just two examples, make it difficult to doubt that at least some kinds of sets affect perception directly.

Therefore, this review must conclude inconclusively with respect to a choice between the two hypotheses. Some evidence exists to support each of them, and some exists which favors one over the other. But there is none that supports one while disproving the other. The issue is still crucial in perceptual theory, and obviously much more very careful research is needed. There is adequate demonstration now of the effects of set on both perception and responses. It still remains to show the conditions under which each will occur.

REFERENCES

AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychological Monographs*, 1951, **65**(7, Whole No. 324).

AVERBACH, E., & CORIELL, E. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, **40**, 309–328.

BORING, E. G. Attribute and sensation. *American Journal of Psychology*, 1924, **35**, 301–304.

BROADBENT, D. E. *Perception and communication.* New York: Pergamon Press, 1958.

BRUNER, J. S. Neural mechanisms in perception. *Psychological Review*, 1957, **64**, 340–358. (a)

BRUNER, J. S. On perceputal readiness. *Psychological Review*, 1957, **64**, 123–204. (b)

BRUNER, J. S., & POSTMAN, L. On the perception of incongruity: A paradigm. *Journal of Personality*, 1949, **18**, 206–223.

BRUNER, J. S., & POTTER, M. C. Interference in visual recognition. *Science*, 1964, **144**(Whole No. 3617), 424–425.

CHAPMAN, D. W. Relative effects of determinate and indeterminant Aufgaben. *American Journal of Psychology*, 1932, **44**, 163–174.

CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75–84.

DEMBER, W. *The psychology of perception.* New York: Holt, 1960.

ERIKSEN, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: University of Nebraska Press, 1958. Pp. 169–226.

ERIKSEN, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, 1960, **67**, 279–300.

ERIKSEN, C. W. Figments, fantasies and follies: A search for the subconscious mind. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962. Pp. 3–26.

ERIKSEN, C. W., & BROWNE, C. T. An experimental and theoretical analysis of perceptual defense. *Journal of Abnormal and Social Psychology*, 1956, **52**, 224–230.

FITTS, P. M. The influence of response coding on performance in motor tasks. In J. Macmillan (Ed.), *Current trends in information theory.* Pittsburgh: University of Pittsburgh Press, 1954. Pp. 169–226.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychological Review*, 1956, **63**, 317–329.

GIBSON, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychological Review*, 1941, **38**, 781–817.

GLANZER, M., & CLARK, W. H. Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 289–299.

GLANZER, M., & CLARK, W. H. The verbal loop hypothesis: Binary numbers. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 301–309.

GREEN, B. F., & ANDERSON, L. K. Color coding in a visual search task. *Journal of Experimental Psychology*, 1956, **51**, 19–24.

GRUBER, H. E., KULKIN, A., & SCHWARTZ, P. L. The effect of exposure time on mnemonic processing in paired-associate learning. Paper read at Eastern Psychological Association, Atlantic City, April 1965.

HABER, R. N. A replication of selective attention and coding in visual perception. *Journal of Experimental Psychology,* 1964, **67**, 402–404. (a)

HABER, R. N. The effects of coding strategy on perceptual memory. *Journal of Experimental Psychology,* 1964, **68**, 257–362. (b)

HABER, R. N. The effect of prior knowledge of the stimulus on word recognition processes. *Journal of Experimental Psychology,* 1965, **69**, 282–286. (a)

HABER, R. N. Limited modification of the trapezoid illusion with experience. *American Journal of Psychology,* 1965, **78**, 651–655. (b)

HABER, R. N., & HABER, R. B. Eidetic imagery: I. Frequency. *Perceptual and Motor Skills,* 1964, **19**, 131–138.

HABER, R. N., & HERSHENSON, M. The effects of repeated brief exposures on the growth of a percept. *Journal of Experimental Psychology,* 1965, **69**, 40–46.

HARRIS, C. S., & HABER, R. N. Selective attention and coding in visual perception. *Journal of Experimental Psychology,* 1963, **65**, 328–333.

HERSHENSON, M., & HABER, R. N. The role of meaning of the perception of briefly exposed words. *Canadian Journal of Psychology,* 1965, **19**, 42–46.

HINTZMAN, D. L. Classification and aural coding in short-term memory. *Psychonomic Science,* 1965, **3**, 161–162.

HOISINGTON, L. B., & SPENCER, C. Specific set and the perception of subliminal material. *American Journal of Psychology,* 1958, **71**, 263–269.

KRULEE, G. K., PODELL, J. E., & RONCO, P. G. Effect of number of alternatives and set on the visual discrimination of numerals. *Journal of Experimental Psychology,* 1954, **48**, 75–80.

KÜLPE, O. Versuche uber Abstraktion. *Berlin International Congress of Experimental Psychology,* 1904, 56–68.

LAWRENCE, D. H., & COLES, G. R. Accuracy of recognition with alternatives before and after the stimulus. *Journal of Experimental Psychology,* 1954, **47**, 208–214.

LAWRENCE, D. H., & LABERGE, D. L. Relationship between recognition accuracy and order of reporting stimulus dimensions. *Journal of Experimental Psychology,* 1956, **51**, 12–18.

LONG, E. R., HENNEMAN, R. H., & GARVEY, W. D. An experimental analysis of set: The role of sense-modality. *American Journal of Psychology,* 1960, **73**, 563–567.

LONG, E. R., REID, L. S., & HENNEMAN, R. H. An experimental analysis of set: Variables influencing the identification of ambiguous visual stimulus-objects. *American Journal of Psychology,* 1960, **73**, 553–562.

MACKWORTH, J. F. The relation between the visual image and post-perceptual immediate memory. *Journal of Verbal Learning and Verbal Behavior,* 1963, **2**, 75–85.

MCKINNEY, J. P. Verbal meaning and perceptual stability. *Canadian Journal of Psychology,* 1966, in press.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psycological Review,* 1956, **63**, 81–97.

MILLER, G. A., GALANTER, E., & PRIBRIM, K. *Plans and the structure of behavior.* New York: McGraw-Hill, 1960.

NEISSER, U. An experimental distinction between perceptual processes and verbal response. *Journal of Experimental Psychology,* 1954, **47**, 399–402.

POLLACK, I. The assimilation of sequentially encoded information. *HumRRO Research Laboratory Memo Report,* 1952, No. 25.

POSTMAN, L. Perception and learning. In S. Koch (Ed.), *Psychology: The study of a science.* Vol. 5. New York: McGraw-Hill, 1963. Pp. 30–113.

POSTMAN, L., & BRUNER, J. S. Multiplicity of set as a determiner of behavior. *Journal of Experimental Psychology,* 1949, **39**, 369–377.

REID, L. S., HENNEMAN, R. H., & LONG, E. R. An experimental analysis of set: The effect of categorical restriction. *American Journal of Psychology,* 1960, **73**, 568–572.

ROSS, S., YARCZOWER, Y., & WILLIAMS, G. M. Recognition thresholds for words as a function of set and similarity. *American Journal of Psychology,* 1956, **69**, 82–86.

RUBIN, E. Bericht uber experimentelle Untersuchungen der Abstraktion. *Zeitschrift für Psychologie,* 1913, **63**, 386–397.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs,* 1960, **74** (No. 11, Whole No. 498).

SPERLING, G. A model for visual memory tasks. *Human Factors,* 1963, **5**, 19–31.

TAUB, H. A. Effects of differential value on recall of visual symbols. *Journal of Experimental Psychology*, 1965, **69**, 135–143.

TEICHNER, W. H., REILLY, R., & SADLER, E. Effects of density on identification and discrimination in visual symbol perception. *Journal of Experimental Psychology*, 1961, **61**, 494–500.

TEICHNER, W. H., & SADLER, E. Effects of exposure time and density on visual symbol identification. *Journal of Experimental Psychology*, 1962, **63**, 376–380.

WILCOCKS, R. W. An examination of Külpe's experiments on abstraction. *American Journal of Psychology*, 1925, **36**, 324–341.

(Received May 21, 1965)

ERRATA

On page 182 of "Toward a Theory of Behavioral Contagion" by Ladd Wheeler in the March 1966 issue, the third sentence in the first paragraph should read: "Thorpe is an ethologist and would probably forgive students of human behavior for insisting that the behavior need not be instinctive."

On page 183 of the same article the third block under the head "Behavioral contagion" should read "Reduction of observer's restraints against B_N. Fear reduction."

On page 190 of this article the first paragraph in the second column should read: "Note that while in the Festinger et al. (1952) formulation deindividuation leads to restraint reduction, the contagion formulation suggests that most of the deindividuation occurs after the restraint reduction."

Psychological Review
1966, Vol. 73, No. 4, 352–356

THEORETICAL NOTES

HABITUATION AND STIMULUS NOVELTY:

A MODEL BASED ON CLASSICAL CONDITIONING

LARRY STEIN[1]
Wyeth Laboratories, Inc., Philadelphia

A theory of habituation is described which assumes that: (a) The arousal reaction is controlled by antagonistic excitatory and inhibitory brain mechanisms; (b) a novel stimulus activates the excitatory mechanism which in turn activates the inhibitory mechanism; and (c) the inhibitory process is conditionable. The inhibitory mechanism becomes conditioned (by Pavlovian conditioning) to the onset of the stimulus itself after repeated presentations. Habituation of arousal is said to occur when the anticipatory, conditioned activation of the inhibitory mechanism overrides the direct activation of the excitatory mechanism. Reactivation of the arousal reaction when the habituated stimulus is changed is explained by a reduction in conditioned inhibitory activity due to stimulus-generalization decrement.

When a signal is presented repetitively to an organism in a well-controlled environment, response to the signal habituates. On initial presentation, the signal elicits an investigatory reaction and other signs of behavioral alerting. Subsequent presentations of the stimulus lead to progressive attenuation of the arousal pattern until it becomes very weak or disappears altogether (Sharpless & Jasper, 1956).

Habituation of the arousal reaction is not due to simple fatiguing of sensory or effector elements. The process clearly does not depend on effector fatigue since a response that has become habituated to a familiar stimulus may still be elicited by a novel stimulus of equal intensity. It is also clear that habituation does not depend on sensory fatigue or nerve accommodation since (*a*) it develops even when long periods of rest intervene between presentations of the stimulus, and (*b*) the orienting response may be restored, after habituation is developed to an intense stimulus, by suddenly reducing the intensity of the stimulus (Sokolov, 1960). Since a mild stimulus could hardly activate a sensory receptor fatigued by strong stimulation, Sokolov's experiment plainly shows that the arousal reaction depends on stimulus change or novelty, and that habituation develops as stimuli become familiar. It follows, therefore, that the brain must have some means of distinguishing between novel and familiar stimuli in these situations.

It is commonly supposed that the detection of stimulus novelty requires that some representation of prior stimulation be coded and stored in the brain so that incoming signals can be matched against the stored signals for similarity. If an incoming signal matches one of the stored signals, the arousal response is suppressed; if no match occurs, the arousal response is evoked (Sokolov, 1960). Although it is conceivable that the brain can in fact store all prior input, scan the store with each incoming signal, and perform the required matching operation, still it seems that such a hypothesis assumes as much as it explains. Other assumptions will be required as well. For example, an additional mechanism would be needed to classify and index the store of coded signals. Otherwise, the time required to scan the store would increase as the store is enlarged by experience, and one

---

[1] The helpful comments of M. E. Bitterman, P. L. Carlton, and R. F. Thompson are gratefully acknowledged.

is led to the improbable conclusion that detection of stimulus novelty should take longer as organisms grow older.

In any event, it is the aim of the present paper to suggest an alternative mechanism for the detection of stimulus novelty in habituation experiments that does not require the brain to perform complex scanning, matching, and indexing operations. The present hypothesis, based entirely on classical conditioning principles, derives chiefly from the suggestive comments of Pavlov (1927, p. 256) on the disappearance of the investigatory reflex, but also contains elements from other theories (Sokolov, 1960; Thorpe, 1956).

Following Sokolov (1960), it is assumed that two antagonistic neural systems are sequentially activated by the initial presentation of a stimulus. Activation of the first system excites or facilitates the arousal reaction, whereas activation of the second system inhibits the arousal reaction. One may assume that the excitatory system is activated by the onset of the novel stimulus and the inhibitory system by its prolongation (Sokolov, 1960); or better, one may assume that the onset of the stimulus activates the excitatory system, which in turn engages the inhibitory system. This sort of negative-feedback arrangement resembles in formal characteristics, though not in time course, the recurrent inhibitory pathways of Eccles (1964, pp. 209–215) and would serve to decrease the sensitivity of the arousal mechanism for a period of time after it is activated. For this reason, perhaps, the arousal reaction habituates more rapidly if presentations of the signal are massed and tends to recover over time (spontaneous recovery—Thompson & Spencer, 1966). In any case, the discussion below assumes, for the sake of simplicity, that sufficient time elapses between trials to allow this residual inhibitory activity to subside.

It is further assumed that the activity of the inhibitory system is conditionable, in the Pavlovian sense, to the onset of the signal. If so, on the next presentation, the signal will activate both neural systems simultaneously rather than sequentially—the excitatory system directly as before[2] and the inhibitory system anticipatorily by conditioning. The excitatory system will in most cases be activated more vigorously than the inhibitory system, since only one trial of inhibitory conditioning will have preceded the presentation of the stimulus. The arousal reaction therefore will again be evoked, but its strength will be somewhat diminished. However, the second presentation of the stimulus will constitute another inhibitory conditioning trial. Hence, on the third presentation of the stimulus, the inhibitory system will be activated still more rapidly and vigorously, and the resulting arousal reaction will be correspondingly weaker. Further presentations of the signal will continue the process until the conditioned activity of the inhibitory system finally overrides the directly evoked activity of the excitatory system. At this point, the arousal reaction will no longer occur at all and will be said to be habituated. A diagram of these processes is given in Figure 1.

What if a new stimulus is presented instead of the habituated one? The excitatory system will be directly activated as before, but the inhibitory system will not be activated anticipatorily since conditioning to the new stimulus has not taken place. The new stimulus will therefore evoke an arousal reaction.

If the new stimulus is similar to the habituated stimulus, the inhibitory system may be partially activated as a result of stimulus generalization. In such a case, the probability or vigor of the arousal reaction will be inversely related to the degree of similarity between the stimuli, as Apelbaum et al. (1960) already have reported. The important point to be noted, however, is that any change in the habituated stimulus (including a decrease in its intensity, as in Sokolov's experiment) can produce a

---

[2] It is simplest to assume that the excitatory system is activated in the same strength each time the signal is presented, regardless of the degree of habituation of the arousal response. If decrement is assumed, the theory becomes unwieldy. Excitatory activity will depend, however, on the intensity and other characteristics of the stimulus.

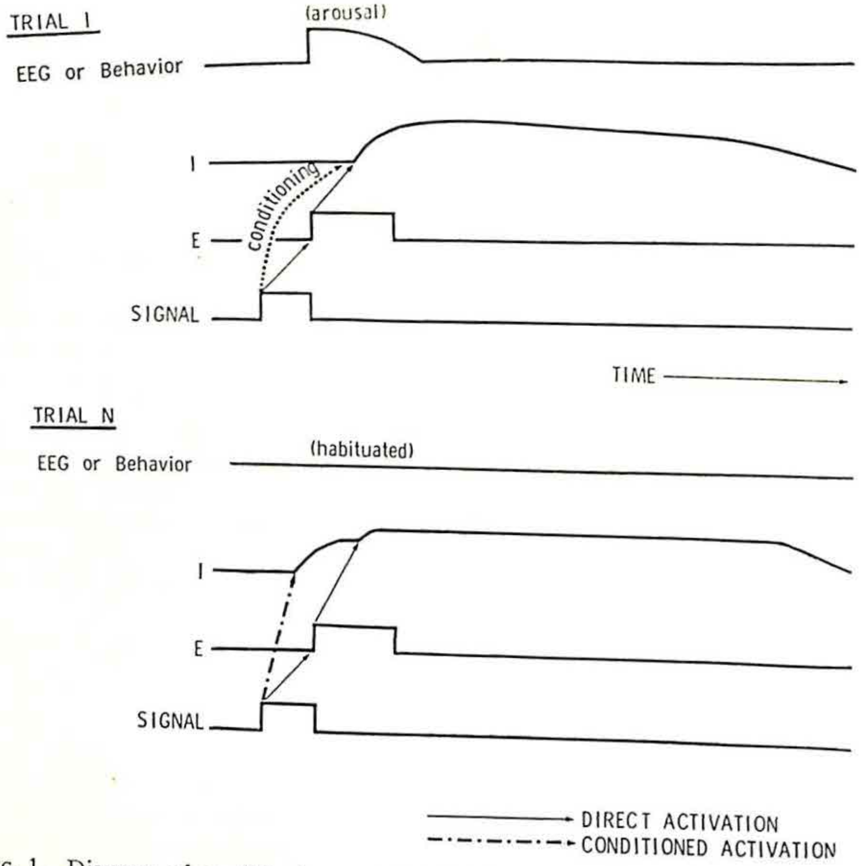

FIG. 1. Diagram of conditioning model of habituation. (TRIAL 1: Signal directly activates excitatory brain mechanism, E, which in turn directly activates inhibitory mechanism, I. Arousal response is elicited while excitatory activity predominates and declines in strength as inhibitory activity grows; furthermore, because inhibition outlasts excitation, the threshold of the arousal response will be elevated for a short time. Since signal onset just precedes activation of inhibitory mechanism, conditioning occurs. TRIAL N: Conditioned activation of inhibitory mechanism to onset of signal overrides direct activation of excitatory mechanism; hence, arousal response fails to occur and is said to be habituated. Further inhibitory conditioning also takes place on this trial despite absence of arousal response, producing below-zero habituation.)

stimulus-generalization decrement in conditioned inhibitory activity and therefore permit the arousal reaction to reappear.

According to this theory, then, stimuli in habituation experiments are familiar to the extent that they have been conditioned to a mechanism that inhibits the arousal reaction. Completely novel stimuli have no conditioned tendency to activate the inhibitory system, whereas stimuli that are very familiar can activate the inhibitory system so rapidly and in such strength that no arousal reaction takes place at all. One should not assume, however, that stimulus familiarity depends exclusively on inhibitory conditioning of the sort described above. For example, if a signal is paired with a primary reinforcer, the signal may acquire familiarity by being conditioned to a neural reinforcing system. The general rule, perhaps, is that stimuli seem familiar to the extent that they can evoke conditioned reactions.

Thompson and Spencer (1966) in a recent review of the habituation literature have detailed the chief habituation phenomena. Their summary statement (pp. 18–20) provides a convenient test of the organizing power of the theory. Some features (e.g., spontaneous recovery, stimulus generalization) have already been discussed; others are treated below.

One key finding, "ubiquitous as habituation itself" (p. 19), is that dishabituation occurs when an additional (usually strong) stimulus accompanies the habituating stimulus. According to the present theory, such dishabituation would happen for two reasons. First, direct activation of the excitatory mechanism would increase because of the additional input provided by the dishabituating stimulus. And second, conditioned activation of the inhibitory mechanism would decrease if it is correct to assume that a stimulus complex consisting of the original stimulus plus the additional stimulus differs enough from the original stimulus alone to produce a stimulus-generalization decrement. Upon repeated application of the additional stimulus, of course, habituation to the stimulus complex would develop in the usual manner.

Another finding that has resisted explanation is the observation that additional habituation trials given after the arousal response has disappeared will result in further response decrement (as indicated by slower recovery or some such measure). Such "below-zero" habituation is puzzling because it suggests that elicitation of the arousal response is not essential for the development, or continued development, of habituation. A strong feature of the present theory is that it is able to handle this phenomenon. The critical assumption is that the excitatory mechanism, and not the arousal response, serves as the "unconditioned stimulus" for the inhibitory process thought to underlie habituation. And although the arousal response diminishes in strength and eventually habituates, the excitatory mechanism does not (see Footnote 2). This ensures an adequate unconditioned stimulus for inhibitory conditioning on every trial, regardless of the degree of habituation. Since inhibitory conditioning can proceed even after the arousal response disappears, continued presentation of signals can produce below-zero habituation.

Finally, what may be said about possible anatomical and biochemical substrates of these hypothetical arousal-regulating mechanisms? There is now general agreement that the reticular formation of the brain stem is the focus of an excitatory mechanism for arousal, and that this mechanism is under adrenergic control (Magoun, 1963b). The focus of the inhibitory mechanism is much more uncertain. It could conveniently be located in the reticular formation itself (Moruzzi, 1964), but other sites are possible as well (Clemente & Sterman, 1963; Magoun, 1963a). Like the inhibitory Renshaw cell arrangement discussed by Eccles (1964), the inhibitory mechanism could be under cholinergic control (Hernández-Peón & Chávez Ibarra, 1963). If so, then it should be possible to retard or prevent habituation by blocking this mechanism by administration of anticholinergic drugs. Such a result has been reported by Carlton (1964, 1965), who found, in a variety of well-controlled tests, that the interfering effect of novel stimulation on behavior does not habituate if familiarization training is carried out under the anticholinergic drug, scopolamine.

REFERENCES

APELBAUM, J., SILVA, E. E., FRICK, O., & SEGUNDO, J. P. Specificity and biasing of arousal reaction habituation. *Electroencephalography and Clinical Neurophysiology*, 1960, **12**, 829–840.

CARLTON, P. L. Brain acetylcholine, nonreward, and habituation. In M. R. Rosenzweig (Chm.), Central cholinergic mechanisms and behavior. Symposium presented at American Psychological Association, Los Angeles, September 1964.

CARLTON, P. L., & VOGEL, J. R. Studies of the amnesic properties of scopolamine. *Psychonomic Science*, 1965, **3**, 261–262.

CLEMENTE, C. D., & STERMAN, M. B. Cortical synchronization and sleep patterns in acute restrained and chronic behaving cats induced by basal forebrain stimulation.

Electroencephalography and Clinical Neurophysiology, 1963, Suppl. 24, 172–187.

ECCLES, J. C. *The physiology of synapses.* New York: Academic Press, 1964.

HERNÁNDEZ-PEÓN, R., & CHÁVEZ IBARRA, G. Sleep induced by electrical or chemical stimulation of the forebrain. *Electroencephalography and Clinical Neurophysiology*, 1963, Suppl. 24, 188–198.

MAGOUN, H. W. Central neural inhibition. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1963.* Lincoln: University of Nebraska Press, 1963. Pp. 161–193. (a)

MAGOUN, H. W. *The waking brain* (2nd ed.) Springfield, Ill.: Charles C Thomas, 1963. (b)

MORUZZI, G. Reticular influences on the EEG. *Electroencephalography and Clinical Neurophysiology*, 1964, **16**, 2–17.

PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford University Press, 1927.

PRIBRAM, K. H. Reinforcement revisited: A structural view. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1963.* Lincoln: University of Nebraska Press, 1963. Pp. 113–159.

SHARPLESS, S., & JASPER, H. Habituation of the arousal reaction. *Brain*, 1956, **79**, 655–680.

SOKOLOV, E. N. Neuronal models and the orienting reflex. In M. A. B. Brazier (Ed.), *Conference on the central nervous system and behavior: III.* New York: Josiah Macy, Jr., Foundation, 1960. Pp. 187–276.

THOMPSON, R. F., & SPENCER, W. A. Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review*, 1966, **73**, 16–43.

THORPE, W. H. *Learning and instinct in animals.* Cambridge: Harvard University Press, 1956.

(Early publication received
November 1, 1965)

Psychological Review
1966, Vol. 73, No. 4, 357–364

STATISTICAL LEARNING MODELS AND INDIVIDUAL DIFFERENCES

R. A. WEITZMAN

Minnesota National Laboratory and the University of Minnesota

In most applications of statistical learning models, it is tacitly assumed, with but little evidence, that individuals do not differ with respect to the models' parameters. This paper examines the evidence, develops a statistical test for the assumption, and applies the test to models and data reported in the literature; the mixed results suggest that individual differences which affect data may not affect the parameters of all models which provide accurate descriptions of the data.

Statistical learning theorists have generally assumed that all the subjects of an experiment for which a model is appropriate have equal values of each of the model's parameters. Although this has been a tacit assumption, it is nonetheless an important one. If generally true, it could obviously have significant implications for the psychology of individual differences. If true but only under special circumstances, it could place serious restrictions on the applicability of statistical learning theory. According to Bush and Mosteller (1955, p. 194), statistical learning theorists "are forced to assume that all subjects have approximately the same parameter values" in order to fit their models to real data. Despite its importance, however, psychologists have devoted little attention to this assumption. Although Audley and Jonckheere (1956) questioned the wisdom of building a statistical learning theory on such an apparently tenuous base, Atkinson, Bower, and Crothers (1965, p. 105) seemed to believe that statistical learning models are likely to be robust with respect to the assumption. Not all assumptions, of course, are directly testable; and, until now, Bower (1962) seems to be the only psychologist to have attempted a direct test of this one. This paper examines Bower's evidence, develops a statistical test for the assumption, applies the test to models and data reported in the literature, and considers possible explanations and implications of the results.

BOWER'S EVIDENCE

Bower (1962) concluded as a result of what he purported to be a statistical test that for his subjects, at least, there were no individual differences with respect to the single parameter c of his model; he based his conclusion on the following premises (see pp. 50–51):

1. The sample variance of total errors made by a subject will tend, under H_0: $c_i = c_j$ for an arbitrary subject j and every subject $i \neq j$, to be equal to, and, under H_1: $c_i \neq c_j$ for an arbitrary subject j and at least one subject $i \neq j$, to be larger than, the theoretical variance for a subject computed by using the formula of the model.

2. Computed values of "F," the ratio of the first to the second of these variances, were not large enough to be significant. (The quotation marks indicate that, according to Bower, this ratio is only approximately distributed as F.)

Although Bower structured his argument so that support of his conclusion depends on the acceptance rather than the rejection of a null hypothesis, still, granting the validity of his premises, his conclusion is not an unreasonable one. If, however, either premise is false, the conclusion remains to be established.

As a matter of fact, neither premise makes sense and so cannot be true or false. With respect to the major premise, it is true that for single-parameter models

TABLE 1

NOTATION

Statistic	Symbol	Formula
Probability of error by an *S* on trial *m*	π_m	Depends on model
Probability of error by an *S* on trials *m* & *M*	π_{mM}	Depends on model
No. *S*s	n	—
Theoretical proportion of *S*s erring on trial *m*	$\bar{\pi}_m$	$n^{-1}\sum_j^n \pi_{mj}$
Theoretical proportion of *S*s erring on trials *m* & *M*	$\bar{\pi}_{mM}$	$n^{-1}\sum_j^n \pi_{mMj}$
Total no. errors for an *S*	T	—
Sample Mean of *T*	$\bar{T}$	$n^{-1}\sum_j^n T_j$
Theoretical variance of *T* overall	$\mathrm{Var}(T)$	$\sum_m \sum_M (\bar{\pi}_{mM} - \bar{\pi}_m\bar{\pi}_M)$
Theoretical variance of *T* for an *S*	$\mathrm{Var}(t)$	$\sum_m \sum_M (\pi_{mM} - \pi_m\pi_M)$
Average theoretical variance of *T* for an *S*	$\overline{\mathrm{Var}(t)}$	$n^{-1}\sum_j^n \mathrm{Var}(t)_j$
Sample estimate (biased) of Var(*T*)	$S_T{}^2$	$n^{-1}\sum_j^n (T_j - \bar{T})^2$
Sample estimate (unbiased) of Var(*T*)	$s_T{}^2$	$n(n-1)^{-1}S_T{}^2$
Model estimate (biased) of Var(*t*)	$S_t{}^2$	Depends on model
Model estimate (unbiased) of Var(*t*)	$s_t{}^2$	Depends on model
Model estimate (unbiased) of $\overline{\mathrm{Var}(t)}$	$\overline{s_t{}^2}$	$n^{-1}\sum_j^n s_{tj}{}^2$

like Bower's the sample variance is under H_1 an overestimate of the *average* theoretical variance for a subject (see Table 1). This fact is implied by Theorem 1 of the Appendix. It cannot be true, however, that under H_1 the sample variance is either an over- or an underestimate of or even equal to the theoretical variance for a subject, because under H_1 there is no single value of the theoretical variance for a subject, subjects with different parameter values having, according to the model, different values of this variance. As for the minor premise, there is no way to compute the appropriate number of degrees of freedom for the denominator of "F." Bower, acknowledging no problem here, evidently assumed this number to be equal to infinity. Such an assumption is certainly unwarranted, however; even under H_0, the value of the theoretical variance is not known, but can only be estimated from the same data as the sample variance. A nonsignificant value of "F," therefore, contributes nothing to the argument which, stripped of its constituent premises, is no argument at all.

A TEST OF INDIVIDUAL DIFFERENCES

Although Bower's evidence does not withstand scrutiny, the basic idea of using different estimates of the variance of total errors to test for individual differences remains a good one. Theorems 1 and 2 of the Appendix form the basis of a test of individual differences which involves the use of such estimates. In order to avoid confusion in both the development and application of these theorems, it is necessary to use a rather elaborate notation. This notation appears in Table 1. The test of individual differences can be applied to all statistical learning models for which it is possible to compute values of $s_t{}^2$. The following example illustrates the derivation of the formula for this statistic in the context of a specific model.

$s_t{}^2$ *for Theios' Model*

Theios (1963) presented a model similar to Bower's; Theios' model involves a parameter, *c*, which is to be interpreted in the same way as the parameter *c* of Bower's model. Theios, assuming (like Bower) a common value of *c* for all his subjects, estimated this value by $N/\bar{T}$, where N designates the number of transitions between successive learning states which (according to the model) a subject must undergo before achieving complete learning. Substituting the corresponding estimate for a single subject, N/T, into the expression which Theios provided for $\mathrm{Var}(t)$ yields:

$$S_t{}^2 = T(T - N)/N.$$

If the expected value of $S_t{}^2$ is equal to some constant H times $\mathrm{Var}(t)$, then $s_t{}^2 = S_t{}^2/H$; only when $H = 1$ is $s_t{}^2$ equal to $S_t{}^2$. To find the value of H, it is necessary to determine the expected value of $S_t{}^2$. If this expected value is denoted by $E(S_t{}^2)$, then, in the context of Theios' model, d designating the differential operator,

$$\begin{aligned}
E(S_t^2) &= \sum_{T=N} S_t^2 \binom{T-1}{N-1} c^N (1-c)^{T-N} \\
&= \sum_{T=N} T(T-N)(1/N) \binom{T-1}{N-1} c^N (1-c)^{T-N} \\
&= \sum_{T=N} (1/N!)T(T-1)(T-2)\cdots(T-N)c^N(1-c)^{T-N} \\
&= (1/N!)c^N(1-c)d^N(\sum_{T=N} T[1-c]^{T-1})/dc^N \\
&= (1/N!)c^N(1-c)d^N(\sum_{T=1} T[1-c]^{T-1})/dc^N \\
&= (1/N!)c^N(1-c)d^N(1/c^2)/dc^N \\
&= (1/N!)c^N(1-c)(N+1)!/c^{2+N} \\
&= (N+1)(1-c)/c^2,
\end{aligned}$$

the fourth equation from the end being true because the N^{th} derivatives with respect to c of $1, 2(1-c), 3(1-c)^2, \cdots, N(1-c)^{N-1}$ are equal to zero.

Since according to Theios $\mathrm{Var}(t) = N(1-c)/c^2$, $H = (N+1)/N$ and

$$s_t^2 = T(T-N)/(N+1).$$

The preceding example serves not only to illustrate the derivation of a formula for s_t^2, but also to help clarify the distinctive meanings of s_t^2 and s_T^2 (see Table 1).

Test Statistic z_0

Under H_0, according to Theorems 1 and 2, s_T^2 and $\overline{s_t^2}$, being unbiased estimates respectively of $\mathrm{Var}(T)$ and $\overline{\mathrm{Var}(t)}$, should be equal. This being so, the statistic

$$z_0 = \bar{D}\sqrt{n}/s_D,$$

where, with

$$D_j = s_{tj}^2 - n(n-1)^{-1}(T_j - \bar{T})^2,$$

$$\bar{D} = n^{-1} \sum_j^n D_j = \overline{s_t^2} - s_T^2$$

and

$$s_D^2 = (n-1)^{-1} \sum_j^n (D_j - \bar{D})^2,$$

is under H_0 for large n (say, $n > 30$) approximately normally distributed with zero mean and unit variance. For models involving a single parameter, according to Theorem 1, z_0 provides a lower-tailed test of individual differences, a significant value of z_0 indicating that different subjects have different values of this parameter. For models involving two or more parameters, according to Theorem 2, z_0 provides a two-tailed test of individual differences, a significant value of z_0 indicating not only that different subjects have different values of one or more of the parameters, but also, if z_0 is larger than zero, that subjects typically perform on one side of average early and on the other side of average later in learning (see Theorem 2).

APPLICATIONS

Application of the test just developed to the models and data reported by Theios (1963) leads to some interesting results. Theios considered two models, one (Model 1) involving one parameter and the other (Model 2) involving twc parameters. Aside from their different numbers of parameters, however, the two models are identical. The formula for s_t^2 derived in the preceding section but one applies with $N = 2$ to Model 1 and with N considered a parameter, its value to be determined by data, to Model 2. Theios applied Model 1 to the data of rats run by himself in a study of traumatic avoidance learning; considering the fit to be good, he applied the same model to the data of dogs run earlier by Solomon and Wynn (1953) in a similar study. Considering the fit of Model 1 to the data of Solomon and Wynn to be poor, however, Theios applied Model 2 to these data and, determining N to be equal to 4, obtained now what he considered (by current standards) to be a good fit. Following are the results of applying the test of individual differences to Model 1 and the data of Theios (1963, p. 416) and to Model 2 and the data of Solomon and Wynn (Bush & Mosteller, 1955, p. 239).

Model 1 and Theios' Data

Testing Theios' data for individual differences with respect to the parameter of Model 1 yielded a value of z_0 equal to .749. Although this value implies that there are no individual differences, it is of course possible that there are individual differences which are so small that the test has insufficient power to detect them. To investigate this possibility, 20 Monte Carlo experiments were run, each involving 100 subjects sampled from a population of subjects whose parameter values were normally distributed with standard deviation equal to .001 and mean equal to .427 (.427 being the number which Theios obtained as an estimate of the parameter value of each of his subjects). Values of z_0 were computed for all the Monte Carlo experiments, each value being smaller than zero. Then, a value of

$$z_1 = (.749 - \overline{z_0})/\sqrt{1 + 1/20},$$

where $\overline{z_0}$ denotes the average value of z_0 for the 20 Monte Carlo experiments, was computed with the result, $z_1 = 1.771$. The chances that Theios' sample was drawn from a population of subjects having parameter values which are normally distributed with a mean of .427 and a standard deviation of .001 or larger are thus less than 5 in 100. This result implies that if there are individual differences among Theios' rats the distribution of these differences, if normal, has a standard deviation smaller than .001. From the standpoint of the psychology of individual differences, this is a provocative result. It means that it is possible to build models which, while ascribing no portion of individual variation in behavior to differential genetic endowment or preexperimental experience, nonetheless are able to describe behavior extremely well (see Theios' Table 2).

Model 2 and the Data of Solomon and Wynn

Testing the data of the dogs for individual differences with respect to the parameters of Model 2 presents the problem of estimating two parameter values from the data of a single subject. Theios, assuming that all his subjects had the same value of each parameter, solved simultaneous equations involving $\bar{T}$ and $S_T{}^2$ to obtain the two parameter estimates. Since for a single subject $\bar{T}$ is equal to T but $S_T{}^2$ is only trivally equal to zero, a substitute for $S_T{}^2$ must be found. Aside from the formulas for the probability distribution and for the mean and variance of total errors, the only formula which Theios presented for Model 2 is the formula for the expected value of the trial number of the last error. This number, $E(L)$, can be estimated from the data of a single subject (fortunately) by the trial number L of the last error. According to Theios,

$$E(L) = (N/c) \sum_{k=1}^{N} (N - k + 1)^{-1}.$$

Substitution in this formula of L for $E(L)$ and (since for Model 1, with N known, c is estimated by N/T) N/T for c results in the equation

$$L/T = \sum_{k=1}^{N} (N - k + 1)^{-1}.$$

Using this equation, N was estimated as follows. Starting with the value 1, successive integral values of N were tried, the estimate of N—$\tilde{N}$—being taken as equal to that integral value for which, with the corresponding estimate of c no smaller than zero or larger than 1, the difference between the two sides of the equation was smallest. Then, c was estimated by $\bar{N}/T$.

With the two parameters estimated for each subject, the value of z_0 was computed, the result being $z_0 = 4.046$. This result not only implies the existence of individual differences, but also, because z_0 is greater than zero, carries the interesting further implication that dogs who made a greater than average number of errors during early trials typically made a smaller than average number of errors during later trials and vice versa. That the dogs actually behaved in this way receives support from two sources. One is the observation of Solomon and Wynn (1953, p. 13) that the correlation between the number of responses preceding the first correct response and the number of alternations between incorrect and correct responses is $-.42$. The other is a correlation of .903 between the individual estimates of N and c. In terms of Theios' model, subjects who proceed rapidly from state to state (have large values of c) must traverse many states (have large values of N) before completion of learning, and vice versa. Before speculating on how widespread such a phenomenon may be, however, it is wise to consider the results in the next section.

"*Model 1-X*" *and the Solomon and Wynn Data*

As pointed out by Theios, there is at least one important difference between his experiment and the experiment of Solomon and Wynn (1953). Although both the rats and the dogs were run in shuttle boxes, the dogs proceeded on alternate trials in different directions, whereas the rats proceeded on every trial in the same direction. Analyzing the Solomon and Wynn experiment as though it consisted on odd trials of one and on even trials of another Theios experiment, Model 1 applying to each of the Theios experiments, leads to a new model for the Solomon and Wynn data. For this model (Model 1-X), instead of the three states—0, 1, and 2—corresponding to the two transitions of Model 1, there are five states, one for each of the possible values of $i + j$ $(i, j = 0, 1, 2)$, a subject in state ij being with respect to Model 1 in state i on odd trials and state j on even trials. According to Theios, the probabilities of responding correctly in successive states are, for Model 1, .00, .50, and 1.00 and, for Model 2 (with $N = 4$), .00, .25, .50, .75, and 1.00, respectively; since for Model 1-X the corresponding probabilities are

$$(.00 + .00)/2 = .00,$$

$$(.00 + .50)/2 = (.50 + .00)/2 = .25,$$

$$(.00 + 1.00(/2 = (.50 + .50)/2$$
$$= (1.00 + .00)/2 = .50,$$

$$(.50 + 1.00)/2 = (1.00 + .50)/2 = .75,$$

and

$$(1.00 + 1.00)/2 = 1.00,$$

Model 1-X and Model 2 (with $N = 4$) should have identical fits to data.

Testing the Solomon and Wynn data for individual differences with respect to the single parameter of Model 1-X resulted in a value of z_0 equal to .530, a value not at all inconsistent with the hypothesis of no individual differences. This, from the standpoint of the psychology of individual differences, is perhaps even more provocative than the result for Model 1 and Theios' data, a result which might, but for a circumstance to be pointed out in the next paragraph, be attributed to the homogeneity of breeding and preexperimental environment characteristic of laboratory rats. The Solomon and Wynn subjects, however, were *mongrel* dogs.

To the extent that Models 1 and 1-X fit the data well and properly, as indeed at least Model 1 seems to do, both the Theios rats and the Solomon and Wynn dogs demonstrated a rather curious phenomenon: They did not by any noticeable amount, positive or negative, transfer the habit of running in one direction to the habit of running in the other. Half of the rats' data represent a first reversal of learning represented by the other half, the earlier learning thus constituting a highly relevant and differentiating preexperimental experience; yet the fit of Model 1 to all these data showed not the slightest tendency for individual parameter values to vary. As for the dogs, Model 1-X explicitly excludes the possibility of transfer between the two habits.

The results of this and the preceding two sections run counter to expectation; it is instructive to attempt an explanation of these results. Although Models 1-X and 2 were fitted to the same data, individual differences affected the fit of Model 2, but not of Model 1-X. Had individual differences failed to show up in the fit of Model 2, the explanation might have been simply that traumatic avoidance learning is too rapid or that data of ones and zeros are too coarse to reveal the existence of individual differences. Because individual differences showed up in the fit of Model 2, however, the explanation must lie elsewhere. In fact, the correlation of .903 between individual estimates of N and c found in the application of Model 2 to the data of Solomon and Wynn provides a clue. If a subject who proceeds from state to state faster than another subject must traverse proportionately more states than the other subject before completion of learning, then the overall learning rates of the two subjects must be equal. Since for Model 1 or 1-X c, being inversely related to the expected number of total errors for a subject, represents overall learning rate, this statement is equivalent to saying that though values of N and c for Model 2 may vary from subject to subject or occasion to occasion the value of c for Model 1 or 1-X must not. In traumatic avoidance learning, individual differences seem to affect the distribution, but not the number of incorrect responses over trials.

A NOTE ON COMPARING THE FITS OF MODELS

Model 1-X with one parameter fits the same data as well as Model 2 with two parameters. This does not necessarily mean that Model 1-X is better than Model 2. The comparison is not fair; Model 2 was fitted without taking into account the effect of individual differences on its parameters. Whereas Model 1-X was fitted properly, Model 2 was not. A proper fit of Model 2 would have to take into account the effect of individual differences. That Model 2 without accounting for this effect fits the data as well as it does, moreover, would seem to be a matter of chance. Had Theios used $E(L)$ instead of $\mathrm{Var}(t)$ for parameter estimation, he would have obtained estimates of N and c equal to 2 and .256, respectively, and the correspondingly poor fit of the dashed line of his Figure 6.

Generally, fairness in comparing the fits of different models to the same data requires testing the fits for individual differences; for a test result which indicates that individual differences affect a model's parameters implies that the fit tested would be better if it took this effect into account than if it failed to do so. The following example illustrates that the improvement in fit might be substantial.

Consider two Model 2 estimates of $\mathrm{Var}(T)$, one which takes into account the effect of individual differences and one which does not. Consider the second estimate first. If Theios had used $E(L)$ instead of $\mathrm{Var}(t)$ for parameter estimation, he would have obtained 22.62 as such an estimate. The corresponding standard deviation is 4.76.

Now, consider the first estimate. A formula for $\mathrm{Var}(T)$ alternative to that of Table 1 is

$$\mathrm{Var}(T) = \sum_{k=0}^{K} k^2\bar{\lambda}_k - \Big(\sum_{k=0}^{K} k\bar{\lambda}_k\Big)^2,$$

where $K = \infty$ and, with λ_k designating the probability that a subject will make a total of $T = k$ errors before completion of learning,

$$\bar{\lambda}_k = n^{-1}\sum_{j=1}^{n} \lambda_{kj}.$$

Theios (1963, p. 410) presented a formula for the probability that a subject will make a total of $T = k$ errors before completion of learning. Substitution into this formula of the individual estimates of N and c which were used in the test of individual differences for Model 2 and the data of Solomon and Wynn resulted in corresponding individual estimates, $\tilde{\lambda}_{kj}$, of the λ_k; substitution of the $\tilde{\lambda}_{kj}$ (by way of the formula for $\bar{\lambda}_k$) together with a value of $K = 50$ into the preceding formula for $\mathrm{Var}(T)$ led to 4.01 as a Model 2 estimate of $\mathrm{Var}(T)$ which takes into account the effect of individual differences. The corresponding standard deviation is 2.00.

Since $S_T{}^2 = 6.36$ ($S_T = 2.52$), a Model 2 estimate of $\mathrm{Var}(T)$ which takes into account the effect of individual differences fits the data substantially better than a Model 2 estimate of $\mathrm{Var}(T)$ which does not take into account the effect of individual differences.

Theios, using current standards, judged his fit of Model 2 to be good; the fact that, accounting for the effect of individual differences, the fit might be better suggests that testing for individual differences is not necessarily counterindicated by a good fit, as, for example, Atkinson et al. (1965, p. 105) seemed to believe.

REFERENCES

ATKINSON, R. C., BOWER, G. H., & CROTHERS, E. J. *An introduction to mathematical learning theory.* New York: Wiley, 1965.

AUDLEY, R. J., & JONCKHEERE, A. R. The statistical analysis of the learning process: II. Stochastic processes and learning behaviour. *British Journal of Statistical Psychology*, 1956, **9**, 87–94.

BOWER, G. H. An association model for response and training variables in paired-associate learning. *Psychological Review*, 1962, **69**, 34–53.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

SOLOMON, R. L., & WYNN, L. C. Traumatic avoidance learning: Acquisition in normal dogs. *Psychological Monographs*, 1953, **67**(4, Whole No. 354).

THEIOS, J. Simple conditioning as two-stage all-or-none learning. *Psychological Review*, 1963, **70**, 403–417.

(Early publication received
November 11, 1965)

APPENDIX

THEOREM 1

Consider a model consisting of only a single parameter c which represents rate of learning. Let c_j designate the value of c for subject j and c_i the value of c for subject i ($i = 1, 2, \cdots, j-1, j+1, \cdots, n$; $j = 1, 2, \cdots, n$); and define H_0: $c_i = c_j$ for an arbitrary subject j and every subject i and H_1: $c_i \neq c_j$ for an arbitrary subject j and at least one subject i. If $\mathrm{Var}(T)$ denotes the overall theoretical variance and $\overline{\mathrm{Var}(t)}$ the average theoretical variance of total errors for a subject, then, under H_0,

$$\overline{\mathrm{Var}(t)} = \mathrm{Var}(T)$$

and, under H_1,

$$\overline{\mathrm{Var}(t)} < \mathrm{Var}(T).$$

Proof. Under either H_0 or H_1, let $\bar{\pi}_m$ designate the probability of selecting at random a subject who will make an error on trial m, and let $\bar{\pi}_{mM}$ designate the probability of selecting at random a subject who will make an error on both trials m and M. Also, under H_0 or H_1, let π_m designate the probability that a selected subject will make an error on trial m, and let π_{mM} designate the probability that a selected subject will make an error on both trials m and M. Then, under H_0, $\bar{\pi}_m = \pi_m$ and $\bar{\pi}_{mM} = \pi_{mM}$; and, under H_1,

$$\bar{\pi}_m = (n^{-1}) \sum_j \pi_{mj}$$

and

$$\bar{\pi}_{mM} = (n^{-1}) \sum_j \pi_{mMj}.$$

By definition,

$$\overline{\mathrm{Var}(t)} = (n^{-1}) \sum_j \sum_m \sum_M (\pi_{mMj} - \pi_{mj}\pi_{Mj}),$$

which, on rearrangement of terms and appropriate substitution, may be written,

$$\overline{\mathrm{Var}(t)} = \sum_m \sum_M \bar{\pi}_{mM} - \sum_m \sum_M (n^{-1}) \sum_j \pi_{mj}\pi_{Mj}.$$

Define δ_{mj} so that $\pi_{mj} = \bar{\pi}_m + \delta_{mj}$. Then, it is seen that under either H_0 or H_1

$$(n^{-1}) \sum_j \delta_{mj} = 0$$

and that, since π_{mj} is larger or smaller than average as c_j is smaller or larger than average for every trial m, under H_0 $\mathrm{Cov}(\delta_m, \delta_M) = 0$ and under H_1 $\mathrm{Cov}(\delta_m, \delta_M) > 0$ for every pair of trials m and M.

Substituting appropriately in the latter expression for $\overline{\mathrm{Var}(t)}$ and simplifying yield:

$$\overline{\mathrm{Var}(t)} = \Big[\sum_m \sum_M (\bar{\bar{\pi}}_{mM} - \bar{\bar{\pi}}_m \bar{\bar{\pi}}_M)\Big] - \sum_m \sum_M \mathrm{Cov}(\delta_m, \delta_M).$$

Since the expression in brackets is, by definition, equal to $\mathrm{Var}(T)$ and since under H_0

$$\sum_m \sum_M \mathrm{Cov}(\delta_m, \delta_M) = 0$$

and under H_1

$$\sum_m \sum_M \mathrm{Cov}(\delta_m, \delta_M) > 0,$$

then, under H_0,

$$\overline{\mathrm{Var}(t)} = \mathrm{Var}(T)$$

and, under H_1,

$$\overline{\mathrm{Var}(t)} < \mathrm{Var}(T).$$

THEOREM 2

Consider a model consisting of two or more parameters $c, d, \cdots$ which represent learning rate and possibly initial or terminal performance. Let $c_j, d_j, \cdots$ designate the values of $c, d, \cdots$ for subject j and $c_i, d_i, \cdots$ the values of $c, d, \cdots$ for subject i $(i = 1, 2, \cdots, (j-1), (j+1), \cdots, n; j = 1, 2, \cdots, n)$, respectively; and define H_0: $c_i = c_j$ *and* $d_i = d_j$ *and* $\cdots$ for an arbitrary subject j and every subject i and H_1: $c_i \neq c_j$ *or* $d_i \neq d_j$ *or* $\cdots$ for an arbitrary subject j and at least one subject i. If $\mathrm{Var}(T)$ denotes the overall theoretical variance and $\overline{\mathrm{Var}(t)}$ the average theoretical variance of total errors for a subject, then, under H_0,

$$\overline{\mathrm{Var}(t)} = \mathrm{Var}(T)$$

and, under H_1, there being two cases, (a) when performance is consistently above or below average throughout learning,

$$\overline{\mathrm{Var}(t)} < \mathrm{Var}(T)$$

or, (b) when performance is on one side of average early and on the other side of average later in learning,

$$\overline{\mathrm{Var}(t)} = \mathrm{Var}(T) + W,$$

where

$$|W| \geq 0.$$

Proof. Proceed as in the proof of Theorem 1, establishing the equation

$$\mathrm{Var}(t) = \Big[\sum_m \sum_M (\bar{\bar{\pi}}_{mM} - \bar{\bar{\pi}}_m \bar{\bar{\pi}}_M)\Big] - \sum_m \sum_M \mathrm{Cov}(\delta_m, \delta_M).$$

Since the expression in brackets is, by definition, equal to $\mathrm{Var}(T)$ and since, under H_0,

$$\sum_m \sum_M \mathrm{Cov}(\delta_m, \delta_M) = 0;$$

under H_1 (case a),

$$\sum_m \sum_M \mathrm{Cov}(\delta_m, \delta_M) > 0;$$

and under H_1 (case b), $\mathrm{Cov}(\delta_m, \delta_M)$ being greater or less than zero for different pairs of values of m and M,

$$\Big|\sum_m \sum_M \mathrm{Cov}(\delta_m, \delta_M)\Big| \geq 0;$$

then, under H_0,

$$\overline{\mathrm{Var}(t)} = \mathrm{Var}(T);$$

under H_1 (case a),

$$\overline{\mathrm{Var}(t)} < \mathrm{Var}(T);$$

and, under H_1 (case b),

$$\overline{\mathrm{Var}(t)} = \mathrm{Var}(T) + W,$$

where

$$|W| \geq 0.$$

Psychological Review
1966, Vol. 73, No. 4, 365–371

OPTICAL TEXTURE VERSUS RETINAL PERSPECTIVE:

A REPLY TO FLOCK[1]

ROBERT B. FREEMAN, JR.
Pennsylvania State University

Differences in numerical analysis and experimental predictions resulting from perspective theory and texture-gradient theory are discussed. Flock's criticisms of the mathematical specification of perspective are shown to be in error. Conversely, Flock's use of regression coefficients as a measure of perceived slant is criticized as being inadequate for description of the data. Finally, results of experiments showing the effects of texture gradients on visual slant are analyzed in terms of perspective theory. By means of simple assumptions, the data of Flock and others can be accounted for better, at least in a qualitative sense, in terms of perspective theory than in terms of optical theta.

This reply to Flock's (1965) note is a discussion of Flock's numerical analysis of the results of investigations of visual slant, of his unfounded criticisms of my specification of retinal perspective, of the distinction between his theory of the stimulus and my theory of perception, and of alternative variables for the specification of texture perspective.

NUMERICAL ANALYSIS

Recent empirical studies, cited in Freeman (1965) and Flock (1965), have demonstrated the relative ineffectiveness of texture gradients as a cue to slant, when compared with the effects of linear retinal perspective. Flock explains these findings by pointing out that the difference in visual angle size of the nearest and farthest "texture element" of surface textures even at the most extreme slants is very small, and probably not discriminable in most cases. Flock is certainly correct, since typical texture elements (e.g., small circles scattered about a homogeneous surface) subtend small visual angles, and therefore are capable of generating only minimal size perspective, even for extreme slants. However, Flock's analysis here is irrelevant. Optical texture-gradient theory (Flock, 1964) does not suggest that relative angular size of texture elements might be a cue to slant, although Flock (1965, p. 508) suggests that it does:

> . . . not only did I discuss the question of surface texture and visual acuity, but I also predicted . . . that as a term like $\delta_N - \delta_F$ approaches liminal values, surface texture will be increasingly *in*effectual in eliciting good slant judgments (Flock, 1964a, p. 386).

Flock has misquoted himself. What Flock (1964, p. 386) had suggested was that visual angle ratios of optical *n* (which is his measure of texture perspective) may be very close to unity when the field of view is very restricted.

On this point, I heartily agree with Flock. As will be shown below, linear perspective, as defined in Equation 1, increases with increasing horizontal and vertical separation of perspective-producing contours. To the extent that texture perspective is an effective cue for slant, it also will be much more effective when it is viewed over wide extents of the visual field. The effect of variation of angular extent of the field of view on the perceived slant of a textured surface has two implications for the theory of "optical texture": (*a*) Texture perspective cannot be invariant with variation in field of view, as predicted by Flock (1964, pp. 388–389); (*b*) since perceived slant varies with field of view, optical *theta* cannot "specify accurately the slant of a

---

[1] This note was written in connection with research supported by Grant MH10691-01, National Institutes of Health, United States Public Health Service.

distal surface from optical variables [Flock, 1964, p. 384]."

This assumption is borne out by the study by Flock and Graves reported by Flock (1965, pp. 511–512) in his note. That experiment, which is discussed further below, shows that the effectiveness of a random texture varies widely with variation in field of view. This finding infirms the theoretical value of optical *theta,* and shows the expected nonveridicality of slant judgments dependent upon texture cues.

Flock then states that the data in Freeman's (1966b) investigation of visual slant using textureless rectangles are "artifacts of his experimental method" (Flock, 1965, p. 509) because of the matching technique used. He says that "*S*s' tasks as they accepted them may have been to match trapezoids for similarity, rather than slanted rectangles for slant equality [pp. 509–510]." He then goes on to point out that, even under circumstances where linear perspective was presumably subliminal in my experiments, the regressions of matched slant on the physical slant of the various standard rectangles were generally close to 1.0. There are two comments which I should like to make concerning Flock's statements. In the first place, if linear perspective had been irrelevant as a cue to slant, subjects would not have been affected to the extent that they were by perspective in their slant-matching task. Their responses would have been more nearly accurate, regardless of rectangle size, had they matched visual angle axis ratios rather than attempting to match perspective. In fact, as Flock points out, when perspective was subliminal, slant matches were quite unaffected by physical (and angular) size, indicating that perspective is a dominant cue when it is discriminable. Second, the use of a regression coefficient here is inappropriate and without descriptive value. Because of the very wide range of stimulus slants used in my experiment, experimental variance of slant matches attributable to the effects of perspective will be small relative to the variance of slant settings of the standard. Consequently, the magnitude of the obtained regression coefficients depended as much on the range of slant settings used as it did on the experimental conditions under which they were viewed. This objection applies to the use of regression coefficients in the description of all of the data referenced by Flock in Table 1 (Flock, 1965, pp. 506–507) and in the text. The individual plotting of response measures (data points) is a more desirable method for the presentation of data and will doubtless continue in use by investigators of perceptual psychophysics.

RETINAL PERSPECTIVE AS A CUE FOR SLANT

Retinal perspective may be specified in various different ways, but generally refers to the convergence of the retinal separation of the projected locus of points of equal physical separation with increasing distance from the eye. Perspective may then be measured by obtaining the difference in visual angle separation of specified near and far pairs of points. When contour perspective is involved, the difference angle in the projection of the near and far edge may be taken as the measure of perspective. This measure varies not only with the slant of the stimulus surface, but also with its height, width, and distance. For a stimulus rectangle slanted about a horizontal axis perpendicular to the visual axis, perspective may be approximated by the relationship,

$$\delta = \tan^{-1} \frac{2\,ad \sin\theta}{c^2 + d^2 - a^2 \sin^2\theta}, \quad [1]$$

where a is the linear distance from the upper or lower edge of the rectangle to the axis of rotation, d the linear distance from the left or right edge to the vertical center line, c the distance of the center of the rectangle to the observer's eye, and θ the (positive or negative) slant of the stimulus. The equation yields a value, δ, which, for rectangles which are viewed normally to the center point of the axis of rotation, is equal to one-half the visual angle difference between the near and far edge of the slanted rectangle.

In his Footnote 7, Flock (1965, p. 513) charges that this equation is "false." Flock bases his statement on his assumption that, as the observation distance (c) decreases to zero, tan δ, and therefore δ, should also approach zero. Flock's analysis is, of course, incorrect. On the contrary, as c assumes very *large* values, δ, and hence retinal perspective, approaches zero.[2] In fact, the angular separations of *all* pairs of points decrease with increasing observation distance. Flock's next two statements, concerning my equations for a possible alternative measure (π) of perspective, are self-contradictory, as can be shown by quoting them, and therefore require no comment: "His [Freeman's] statement that π is unchanged by the height of the rectangle is self-evident. . . . The angular size of π at the eye very definitely changes with the height of the rectangle."

Finally, Flock (1965, p. 511) charges me with a "reluctance to accept any other variable than the contour convergence of slanted rectangles as a stimulus for perceived slant. . . ." His statement is not consistent with my position that with ". . . monocular observation, visual shape and visual slant are a function *primarily* of linear outline perspective [Freeman, 1965, p. 503; italics added]." When contour perspective and texture perspective are both visible in a monocularly-viewed slanted stimulus, texture gradients may well have a minor effect on perceived slant. However, the effect should be primarily *negative*, if the geometrical analysis in Equation 1 is related to visual slant. Since a and d assume maximal values at the edges of a slanted stimulus, the response of the eye to the retinal perspective (δ) arising from those edges must also be maximal. Conversely, perspective convergence of pairs of points ("texture elements") within the boundaries of the stimulus figure must take smaller values. Thus perceived slant could (although not necessarily must) be decreased as a result of visible texture gradients, but cannot be increased! This will be true even if internal (element) and contour perspective are averaged by the eye, since the perspective difference angles (δs) of a substantial proportion of pairs of element points must generate sensations of minimal slant. Conversely, if contour perspective takes such a low value that it is not discriminable by the eye (as happens at small slants and in very small stimuli at any slant), then texture gradients must also be indiscriminable, since their visual angle variation is necessarily less than that of the bounding contours.

THEORY OF STIMULI VERSUS THEORY OF PERCEPTION

In my earlier note (Freeman, 1965) I criticized optical *theta* as being a theory of stimuli rather than a theory of perception. Flock (1965, p. 514) does not feel that anything is gained by the distinction. On the contrary, a great deal is gained by the distinction, since the success of investigations of perceptual psychophysics may be seriously affected by the selection of physical parameters with which to specify the proximal stimulus to the receptor surface. The proximal stimulus is often specified in terms of some transformation of the energy as measured at the energy-producing source. Like the visual angle itself, optical *theta* is such a transformation. Since "the slant angle, θ, of the surface is given by . . . optical *theta* [Flock, 1964, p. 388]," then perceptions which depend upon optical *theta* must be determined by physical slant, θ. That is, *slant perceptions determined by texture gradients (optical theta) must be veridical.* If this is the case, then optical *theta* may be considered an effective psychophysical stimulus, and the transformations of the surface texture implicit in Flock's Equation 5 indicative of the analyzing mechanisms of the human eye. However, if visual responses are not dependent upon optical *theta*, then texture-gradient theory is a theory of the visual stimulus which is meaningless for the analysis of vision.

---

[2] Equation 1 and other formulas for specifying the visual angle projection of slanted figures are derived in Freeman (1966c, Appendix). For reasons of economy, they are not reproduced here.

ALTERNATIVE VARIABLES OF TEXTURE PERSPECTIVE

A previously unpublished study by Flock and Graves, described briefly by Flock (1965, pp. 511–512), indicates that random-texture gradients, although weak, can function as a cue to slant. Flock (1965, p. 511) asks how linear perspective can predict the findings of this experiment whose slant stimuli were random-texture fields and lacked linear contours. This section will show that: (*a*) The theory of optical *theta* cannot account for even the data of Flock and Graves; and (*b*) four postulates derived from linear-perspective theory are capable of predicting their results in a qualitative sense.

The "elements" of the textured field used by Flock and Graves were large triangles of random size and shape, averaging 4.5° in visual angle extent. Viewed through apertures of 20°, 40°, and 80° of solid angle, at each of six unspecified slants, the regressions of the unspecified slant judgments on physical slants were .10, .30, and .69 respectively. None of the three regression coefficients approached the value of 1.0 predicted by the optical *theta* transformation, and all of them are exceedingly low in view of the very large visual fields used. Finally, the regression coefficients increased with increasing viewing field—a finding not predicted by optical *theta*.

Granting the lack of psychological validity of optical *theta*, how is it possible for the eye to respond to texture-density gradients at all?[3] Stated in other terms, what properties of the visual texture vary sufficiently with variation in slant to be detectable by the eye? I should like to suggest four visual cues which may, under some stimulus conditions, such as those used by Flock, contribute to a perception of slant. None of these requires any unusual assumptions about the discriminative capacities of the visual system.

1. For relatively large texture elements, large visual fields, and sizable slants, the projective difference in visual angle size of near and far texture elements may well contribute to the perception of slant. This is possible under two circumstances: (*a*) when texture elements are all of the same size, in which case the visual task is similar to that of responding to linear perspective, (*b*) when the relative size of irregular texture elements is known or perceivable. In any experiment which uses repeated presentations of a textured surface at a variety of slants, it is of course possible for the observer to obtain information concerning texture element size at the near and far boundaries of the visual display.

The size perspective of equally (or similarly) sized texture elements can be obtained with an equation of the form of Equation 1 when the distance (*a*) of the element from the axis of rotation, the diameter of the element (2*d*), and observation distance (*c*) are known, and the entire surface is always visible. In the Flock and Graves experiment, however, the limits of the area of the surface which was visible through the 20°, 40°, or 80° field stop were dependent upon the slant of the textured surface. Under such conditions, the visual angle difference (2δ) between the elements at the near and far extremes of the visual aperture is given exactly by means of the law of sines:

$$\delta = \tan^{-1} \frac{2\,cd \sin\alpha \tan\theta}{c^2 + d^2(1 - \sin^2\alpha \sec^2\theta)}, \qquad [2]$$

---

[3] In his Footnote 6, Flock (1965, p. 512) explains how linear outline perspective can be considered an "extreme and special case" of optical *theta*. However, the visual response to linear perspective does not conform to optical *theta*, but rather to retinal perspective as defined in Equation 1. Visually matched slant (Freeman, 1966b) increases and slant threshold decreases (Freeman, 1966a) with the increasing linear dimensions of the outline rectangle as predicted by perspective theory.

where c is the distance of the observer's eye from the axis of rotation, $2d$ the diameter of the texture elements, θ the slant of the surface from the frontoparallel plane, 2α the visual angle limit of the visual field, and 2δ the total visual angle difference between the near and far texture elements. Equation 2 is appropriate only when both the near and far elements are centered in the visual field.

One can gain an appreciation of the enormous size perspective possible under the conditions reported by Flock and Graves by means of Equation 2. These investigators report an average visual angle element size at 0° slant of 4° 30′. Let us assume that this angle was produced by a texture element whose diameter was 10.6 centimeters ($d = 5.3$ centimeters) viewed at a distance (c) of 135 centimeters. Let us assume that a surface is viewed at a slant (θ) of only 40° through an aperture of 80° ($\alpha = 40°$). Using Equation 2, we obtain a value of $\delta = 2° 25.5'$. The total visual angle difference is therefore $2\delta = 4° 51'$.

The value of nearly 5 *degrees* of size perspective is clearly large enough to be a strong factor in the visual perception of the slant of textured surfaces. Such a very large value for 2δ is the result of the fact that, when viewed at a substantially large angle through a large aperture, the visible linear extent of a textured surface in depth is also very large. Letting a be the linear distance from the axis of rotation to the farther limits of the visual field, and a' the linear distance from the axis of rotation to the nearer limits, then, by the law of sines,

$$a = \frac{c}{\operatorname{ctn} \alpha \cos \theta - \sin \theta}, \qquad [3a]$$

$$a' = \frac{c}{\operatorname{ctn} \alpha \cos \theta + \sin \theta}. \qquad [3b]$$

Combining and simplifying,

$$a + a' = \frac{2c \sin 2\alpha \cos \theta}{2 \cos^2 \theta + \cos 2\alpha - 1}, \qquad [4]$$

where all values are as defined above. With the values $\alpha = 40°$, $\theta = 40°$, $c = 135$ centimeters, the obtained $(a + a')$ is 586.51 centimeters. This value is more than 4 times the distance from the eye of the observer to the axis of rotation of the textured surface.

2. A second possible cue for slant under such conditions may be monocular accommodation and image clarity (or lack of it). Although these cues have been found to be very weak or even negative (Heinemann, Tulving, & Nachmias, 1959), they may play a minor role in space perception when large distance differences are involved.

3. When the same textured surface is presented repeatedly to the same subject at different slants, the total *number* of visible elements may well be a conditional cue to slant perception when surfaces are viewed through apertures of fixed visual angle as in the Flock and Graves experiment. Assuming equal average separation of texture elements at all points of the visible surface, the number of visible elements must therefore increase as a linear function of increase of visible area of the surface with increasing slant. It can only be a conditional one, however, since the number of visible elements is a function of $(a + a')$, which is equal at equal slants in either direction, multiplied by the density of the elements. If other cues for direction of slant are effective, then number of visible elements can provide a secondary cue to amount of slant.

4. Finally, there may also be some perspective cues generated by the change in average angular separation of texture elements. In texture-gradient theory, the response to texture perspective requires that the eye be able to "count off equal numerical units of congruent luminous patterns along an optic meridian [Flock, 1965, p. 389]." Such an assumed mechanism of vision seems most improbable, especially in view of the demonstrated inability of the eye to estimate quantity beyond a critical value of about seven (Kaufman, Lord, Reese, & Volkmann, 1949; Miller, 1956). A more parsimonious assumption would be that the visual system responds *as if* there were linear

contours whose retinal convergence could then be evaluated. Presumably the eye could accomplish this only if it were capable of ordering the texture elements linearly. This in turn would be possible only when the variability of texture-element separations falls below some critical ratio to the average texture-element separation. The texture-variability assumption would then be the only one necessary to explain the capability of the eye to register texture perspective. Conversely, when the variability of the texture elements *exceeds* that critical value, the eye should be incapable of selecting near and far element pairs of approximately equal linear separation whose visual angle differences are to be evaluated.

All four of these potential cues for visual slant vary with the solid angle size of a fixed viewing aperture. Variables 1, 2, and 4 (Variable 3 is not applicable) are affected by the linear dimensions of slanted surfaces whose view is not obstructed by a field stop. Since all of these variables are so affected, if any combination of them is effective as a cue for visual slant, judged slant will increase (positively or negatively) with increasing viewing area of a slanted surface. Similarly, the regression coefficients used by Flock will increase with increasing viewing area. These predictions are supported by data of Gibson (1950) and Flock and Moscatelli (1964), as well as by those of Flock and Graves. Flock and Moscatelli point out that regression coefficients obtained in their experiment using a 38° field of view were, for both regular and irregular surface textures, greater by .10 than the regression coefficients they computed for Gibson's (1950) data which were collected with a smaller (24°) viewing aperture.

Finally, the prediction of perspective theory that visual slant will increase with increasing viewing area must be compared with the prediction that the effectiveness of optical *n* will be restricted if the field of view is "too small." Under such circumstances "the number of like elements represented in the optical texture may be too few [Flock, 1964, p. 386]." This statement implies, however, a discontinuous scale of discriminability of slant under texture theory, from absence of discriminability under unfavorable conditions (small viewing aperture) to near-perfect discriminability with adequate apertures. It does *not* state that visual slant will vary with aperture size, as predicted by perspective theory, when aperture size has reached or surpassed the critical value. It is this prediction of perspective theory which is supported by the data of Gibson, Flock and Moscatelli, and Flock and Graves.

CONCLUSION

Perspective theory can provide the basis for the prediction of outcomes of other experiments involving texture perspective. Perceived slant of textured surfaces will be correlated with texture-element size, size of visual field, regularity of texture, as well as the physical slant of the surface. Under no circumstances, however, will the eye prove more sensitive to texture perspective than to linear retinal perspective, except when the texture carries a regularity equal to that of outline perspective, as in a checkerboard pattern. Even here, however, the linear perspective generated by the outline of the pattern must produce the greatest perception of slant.

REFERENCES

FLOCK, H. R. A possible optical basis for monocular slant perception. *Psychological Review*, 1964, **71**, 380–391.

FLOCK, H. R. Optical texture and linear perspective as stimuli for slant perception. *Psychological Review*, 1965, **72**, 505–514.

FLOCK, H. R., & MOSCATELLI, A. Variables of surface texture and accuracy of space perceptions. *Perceptual and Motor Skills*, 1964, **19**, 327–334.

FREEMAN, R. B., JR. Ecological optics and visual slant. *Psychological Review*, 1965, **72**, 501–504.

FREEMAN, R. B., JR. Absolute threshold for visual slant: The effect of stimulus size and retinal perspective. *Journal of Experimental Psychology*, 1966, **71**, 170–176. (a)

FREEMAN, R. B., JR. Effect of size on visual slant. *Journal of Experimental Psychology,* 1966, **71**, 96–103. (b)

FREEMAN, R. B., JR. Function of cues in the perceptual learning of visual slant: An experimental and theoretical analysis. *Psychological Monographs,* 1966, **80**(2, Whole No. 610). (c)

GIBSON, J. J. The perception of visual surfaces. *American Journal of Psychology,* 1950, **63**, 367–384.

HEINEMANN, E. G., TULVING, E., & NACHMIAS, J. The effect of oculomotor adjustments on apparent size. *American Journal of Psychology,* 1959, **72**, 32–45.

KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMANN, J. The discrimination of visual number. *American Journal of Psychology,* 1949, **62**, 498–525.

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, **63**, 81–97.

(Early publication received
February 25, 1966)

Psychological Review
1966, Vol. 73, No. 4, 372–375

COMMENTS ON "A POWER COMPARISON OF THE *F* AND *L* TESTS—I"

PAUL A. GAMES

Ohio University

Boersma, DeJonge, and Stellwagen (1964) fail to consider any sitution except that where the specific ordered alternative of the L test is true. They thus fail to consider the general statistical inference consequences of using ordered-hypothesis tests vs. omnibus tests. The power comparison provides little useful information on the parametric vs. nonparametric choice of tests in either the ordered-hypothesis or omnibus test situations. The conclusion of consistently superior power of the L test is somewhat dubious in that the authors failed to demonstrate equal risks of Type I error in the 2 tests and used computer simulation of data that was inappropriate for either the randomized-blocks analysis of variance or the L test.

In their article, "A Power Comparison of the *F* and *L* Tests—I," Boersma, DeJonge, and Stellwagen (1964) seem to disparage the general usage of the omnibus *F* test. They state that ". . . the most frequent practice is to perform the omnibus test first and follow with a directional test only if the omnibus test leads to a rejection of H_o. We do not maintain that this is good practice, but it is almost universal in application [p. 507]." Following this, they state: "This research gives empirical support to the prediction that the *L* test is more powerful for testing ordered hypotheses than the omnibus *F* test for the randomized blocks design [p. 512]." The authors avoid recommending that a potential user drop the *F* test and replace it with the *L* test in his next experiment. Lest the casual reader come to this conclusion after reading the above article, the following points should be considered.

In the *F* test, the null hypothesis, H_o: $M_1 = M_2 \ldots M_k$, is tested against all possible alternatives, or *not* H_o. If H_o is retained, then no further steps are justified, but if H_o is rejected, then further procedures (such as all possible *t* tests) are carried out to determine a likely alternative, such as $M_1 = M_2 < M_3 < M_4$. This procedure certainly leaves room for improvement in that the rejection of the null by the *F* test provides no information on which of the many alternatives should be accepted; any of the subsequent procedures are not well designed for this particular purpose. Other omnibus alternative procedures are being sought that may eventually replace the *F*-*t* sequence. A two-sample sequential procedure is available (Bechhofer, Dunnett, & Sobel, 1954) that enables the experimenter (*E*) to specify: (*a*) a region of relative indifference (e.g., if $|M_i - M_j| \leq s$, the difference is regarded as trivial), and (*b*) the probability (*P*) of arriving at a correct ranking of the *k* population means *if* the adjacent means are more than *s* apart; and then yields a solution for the sample size required to accomplish this task. If *E*s are willing to allow statistical inference procedures rather than administrative considerations to determine sample size, it is possible to arrive at statistical conclusions of greater clarity than: "The population means are not equal," even in the absence of any a priori hypotheses.

However, the *L* test requires that *E* specify a single alternative in which all means involved are a priori ranked. *E* then tests the model of equal population means against this single alternative. If this alternative happens to be true, nothing can be lost by this procedure, and we would usually expect an increase in power with respect to this one alternative condition. When one can specify population means on a computer, and even arrange them in a nice linear trend so the difference between all means is constant, it is hardly surprising that such a one-alternative test has more power than a test which must distribute the probability in the region of rejection over a larger region to detect many other alternatives. Unfortunately, most of the time E's a priori basis for ordering the population means is not on as firm a basis as when using computer simulation. Usually we are doing the experiment and carrying out a statistical analysis precisely because we have ample background to doubt our a priori predictions. Before adopting the L test in such a practical situation, several questions completely ignored by Boersma et al. (1964) are quite important. What are the probabilities of falsely accepting the alternative if another alternative is true? What are the probabilities of falsely rejecting a true null hypothesis (making a Type I error)? What are the consequences and/or the relative importance of the various kinds of error in such situations?

Kaiser (1960, p. 163) has written: "The particularly repugnant . . . 'errors of the third kind' . . . involve deciding upon a difference in the wrong direction." Kaiser was speaking of concluding that $M_1 < M_2$ when in reality $M_1 > M_2$ (in a t-test situation), however, the same kind of error and the same repugnance may be generalized to the k-mean situation. With the L test, it is certainly feasible that one of the many other alternatives to the hypothesis of equal means may be true, but by sampling fluctuation the obtained statistic leads us to reject the null and falsely accept the alternative. If $M_1 = M_2 < M_3 = M_4$ or if $M_1 < M_2 < M_3 > M_4$, the probability of obtaining sample data such that $\overline{X}_1 < \overline{X}_2 < \overline{X}_3 < \overline{X}_4$ and the L test is significant may be high. Other alternatives more drastically different in interpretation might still have an appreciable associated probability of obtaining a significant L statistic. Should the possessor of a significant L statistic take his alternative hypothesis as proven (within the usual limits of statistical inference), or should he have some of the reservations of the usual F-t sequence?

Another untouched statistical inference point in the article is the process to be followed if the null is retained. In the usual F-t sequence this is the end of the game, but what about an E who has predicted $M_1 < M_2 < M_3 < M_4$ and obtained $\overline{X}_1 > \overline{X}_2 > \overline{X}_3 > \overline{X}_4$ or some other interesting result deviant from his prediction? Is he really going to run only the L test against his prediction and stop, or will publication pressure or common sense cause him to resort to conventional methods to test this trend? Running a series of statistical tests on the same data is a sure way to inflate the risk of a Type I error above the nominal value reported (Goldfried, 1959).

E must consider such factors in deciding whether to run an omnibus test or an ordered-hypothesis test. If he has decided to conduct an ordered-hypothesis test, then he has a further choice between the nonparametric L test and the parametric F test for linear trend (Grant, 1956; Winer, 1962). Although the parametric test is sensitive to an additional violation of the null hypothesis that is ignored by the L test (namely $M_1 > M_2 > M_3 > M_4$), these two tests ask the same general question of the data, and a power comparison is pertinent as a basis of choice between them. Similarly, if E has decided on an omnibus test, a power comparison of the parametric F test and the nonparametric Friedman (1937) chi-square test would be valuable to assist in choosing between these two alternative tests. However, the comparison of the power of the L test and omnibus F test fails to clarify the appropriate parametric-nonparametric choice in either situation. For the alternatives used by Boersma et al. it is necessarily true that the F test for linear trend will have greater power than the omnibus F test and that this superiority will increase with k. It thus seems quite likely that if the L test were contrasted with the comparable parametric test, the later would prove to be more powerful. In addition, the following methodological considerations suggest that even the Boersma et al. conclusion of consistently superior power of the L test over the omnibus F test is based upon dubious grounds.

The major concern of statistical inference has traditionally been the control of Type I errors, yet Boersma et al. neglect to establish the relative merits of the tests with respect to the risk of a Type I error. Unfortunately, their study failed to run samples precisely at the point at which $\Sigma(M_j - M_o)^2 = .0$ (the null hypothesis is true), but they did run when $\Sigma(M_j - M_o)^2 = .01$. Under this condition, in Figures 3–6, their data show the *L* test as having probabilities of rejection consistently above the nominal .05 level of significance, while the *F* test stays quite close to this level. In Figures 5 and 6, the probabilities of rejection of the *L* test approximate .15 and .10 at this point. Boersma et al. discuss this as evidence for the superiority of the *L* test for very small differences between the means. An alternative explanation that they do not discuss, and do not eliminate, is that it is evidence for the inferiority of the *L* test in controlling the risk of Type I error. If the *L* test has an actual risk of a Type I error of approximately .10 when using the .05 level of significance, this would be a substantial liability yet would produce elevated power curves. Before attributing such curves to power virtues, it would seem appropriate for the authors to demonstrate that the two tests do indeed have similar levels of risk of Type I errors.

The second objection is that the computer approach employed was completely inappropriate for the two statistics used. Either the randomized-blocks analysis of variance or the *L* test makes sense only in a situation in which we would expect the measures within each block to have "something in common." The model for either design is one in which each observation is assumed to consist of the sum of the grand mean (M_o), a treatment effect (T_i), a block effect (w_j), and a random error (e_{ij}) which includes any block × treatment interaction, or $X_{ij} = M_o + T_i + w_j + e_{ij}$ (Winer, 1962). In the computer operations of the above article, T_i's were added to successive random deviates to produce the various violations of the null; no block constants were added, however. Treatment of the data collected as a $k \times n$ matrix mean that the k random deviates that happened to fall first in each sample were considered as having something in common; the k random deviates that were second in each sample were considered as having something in common, etc. The computer simulation of the data, however, was by procedures in which no "block effects" were present, so the data generated were of the form $X_{ij} = M_o + T_i + e_{ij}$. In this case, the appropriate analysis of variance is the simple randomized independent-groups design, and an appropriate ordered-alternative distribution-free test for comparison would be the Jonckheere (1954) test rather than the *L* test. If a graduate student came with experimental data collected by procedures analogous to those used in the above study and asked which test to run, the randomized blocks or the *L* test, the answer would clearly be neither.

The randomized-blocks design has a major assumption not associated with the independent-groups design, namely, the requirement of equal correlations (across blocks) between all treatments. Boersma et al. have met this assumption, since all correlations obtained in their data are samples from a population with a correlation of zero; it does seem a quite dubious practice however to encourage *E*s to generate random correlations between their observations from independent groups in order to apply either the randomized-blocks design or the *L* test. While such procedures should have no long-run effect on the risk of a Type I error, they do have an adverse effect on the risk of a Type II error. Running a randomized-blocks design on such a set of data results in a loss of power due to the fact that $n - 1$ degrees of freedom are used to assess the nonexistent block effect. This $n - 1$ degrees of freedom is thus lost from the error term and results in an increase in the size of the *F* statistic needed to reject the null hypothesis. For example, in Boersma et al.'s Figure 1, the critical value of *F* for the .05 level would be 4.46 ($df = 2{,}8$), while the critical value in an independent-groups design would be 3.89 ($df = 2{,}12$). The power curve for the correct *F* test would be raised substantially in the middle portions due to a change of this magnitude. The magnitude of change in such comparisons decreases as both k and n increase, but all such power curves would be somewhat higher than the corresponding F curves obtained by Boersma et al.

The use of empirical sampling studies to investigate the characteristics of statistical tests under conditions where mathematical solutions are not yet available seems quite worthwhile, but only when the computer simulation generates the kind of data for which the test is appropriate. Although the present author is sympathetic with Boersma et al. in their desire to reduce the risk of a Type II error, this should not be done by ignoring the risk of Type I errors, nor by ignoring the characteristics of the test when other alternatives than the single a priori prediction are true. Considering the usual limited theory back of most a priori predictions in psychology, the use of the L test seems like shooting at a sparrow on the wing with a rifle. If your aim is precise, a kill is assured, but there is an awful lot of space in which to miss. The shotgun approach of an omnibus test seems more suitable even in situations in which there is some basis for prediction.

REFERENCES

BECHHOFER, R. E., DUNNETT, C. W., & SOBEL, R. A two sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika,* 1954, **41**, 170–176.

BOERSMA, F. J., DEJONGE, J. J., & STELLWAGEN, W. R. A power comparison of the F and L tests—I. *Psychological Review,* 1964, **71**, 505–513.

FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association,* 1937, **32**, 675–701.

GOLDFRIED, M. R. One-tailed tests and "unexpected" results. *Psychological Review,* 1959, **66**, 79–80.

GRANT, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychological Bulletin,* 1956, **53**, 141–154.

JONCKHEERE, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika,* 1954, **41**, 133–145.

KAISER, H. F. Directional statistical decisions. *Psychological Review,* 1960, **67**, 160–167.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received February 23, 1965)

Psychological Review
1966, Vol. 73, No. 4, 376–381

MEDIATION THEORY AND THE "SINGLE-STAGE" S–R MODEL: DIFFERENT?

LEON A. JAKOBOVITS[1]

Center for Comparative Psycholinguistics, University of Illinois

The argument presented in this paper is that the widely held notion of a fundamental difference between so-called single-stage S–R models and 2-stage mediation theories is illusory, and the distinction cannot in fact be maintained in S–R theories of language. Skinner's account of verbal behavior and certain phenomena that relate to it is compared to the treatment given by Osgood, and the basic similarity between them is pointed out.

It is still fashionable today to teach students in various areas in psychology that there are two kinds of S–R models: the so-called single-stage theory vaguely identified with "tough-minded" Skinnerians and the two-stage (or multiple-stage) mediation theory, equally vaguely identified with "tender-minded" cognitive theorists. The purported inadequacy of the single-stage model to handle such ubiquitous phenomena as semantic generalization and symbol manipulation is quickly pointed out, and the single-stage model is then discarded as a useless relic of an earlier and more primitive era in psychology. It is the purpose of this paper to show that the widely held distinction between existing single-stage and mediation theories is illusory and that the explanatory power of the two "types" of theories is actually identical.

THE MEDIATION PARADIGM

There are many specific versions of the mediation paradigm. Perhaps its most general formulation has been recently restated by Goss (1961) as follows:

> When Events A, B, and C occur in the sequences $A \rightarrow B \rightarrow C$ . . . and the probability of the occurrence of C is actually or potentially greater or less when C is preceded by $A \rightarrow B$ than when C is preceded by A alone, B can be described as a mediating process or event [p. 286].

At this level of discourse, the mediation paradigm is neutral with respect to single- or two-stage S–R models. Indeed, as Goss shows, there was "remarkable commonality" in their conceptualization of the mediation paradigm between Associationists (Locke, Mill), Structuralists (Wundt, Titchener), and Early Functionalists (James, Dewey). One might add with confidence here Behaviorists, such as Watson, Lashley, and Hull, as well as contemporary theorists such as Woodworth, Skinner, Hebb, Osgood, and Mowrer.

Disagreements arise as soon as one compares more detailed formulations of the mediation paradigm which attempt to specify the nature of the Events A, B, and C: ideas, emotions, mental states, reactions, neural processes, etc.

OVERT VERSUS COVERT EVENTS?

Explanations of the distinction between single-stage and two-stage S–R models usually come down to the presence or absence of explicit reference to covert, nonobservable events. For example, in one influential version of the mediation hypothesis (Osgood, 1953) the distinction is represented as follows:[2]

---

[1] I am grateful to C. E. Osgood and B. F. Skinner for their criticisms of an earlier draft of this paper. This acknowledgement should not be construed as an indication of either agreement or disagreement on their part with the arguments developed in this final version.

[2] In S–R paradigms such as those below, it is useful to distinguish between learned S–R habits (solid lines) from unlearned R–S or r–s connections (dashed lines) where the

single stage

$$S \xrightarrow{\qquad\qquad} R \quad \text{overt elements}$$

two stage

$$S \rightarrow r_m \dashrightarrow s_m \rightarrow R \quad \text{overt and covert}$$

The covert events $r_m \dashrightarrow s_m$ are said to mediate the overt events $S \rightarrow R$. Note, however, that the single-stage paradigm could easily be rewritten to include mediating events:

$$S \rightarrow R \dashrightarrow S \rightarrow R$$

This is the paradigm for response chaining, and it fits the definition of the general mediation paradigm given above.

In the light of this discussion, it would be incorrect to oppose the mediation hypothesis to single-stage theories. Since there is no theoretical restriction on the number of mediating events, it would also be incorrect to contrast single-stage versus two-stage or multiple-stage theories. It is tempting then to settle for a distinction which would contrast the presence or absence of covert references in the theory. However, it is impossible to find among the known S–R formulations (of the so-called single-stage type) any that do not make an explicit reference to covert responses. Furthermore, there is no logical necessity for mediation theories (of the so-called two-stage type) to insist on the presence of covert events. Theorists like Osgood have emphasized the covert nature of mediating events for two reasons. One relates to the difficulty of explaining certain phenomena, such as semantic generalization, where there are no obvious overt events that can account for the generalization obtained. The second reason relates to the functional advantages for the organism if mediating events can easily become covert. Both Osgood (1953, pp. 397–398) and Skinner (1957, pp. 436–437) have stated these points.

Given the fact that all S–R formulations refer to covert events, the question can be raised (Fodor, 1965) whether these covert events have identical properties in the so-called single-stage and two-stage models. Thus, Fodor argues that the covert mediating events in the single-stage paradigm retain a one-to-one relationship to overt events and hence have no excess meaning and no explanatory power. Indeed it is said that the theoretical usefulness of covert mediating responses in the two-stage paradigm lies in their relative independence from direct environmental stimulation (e.g., as in symbolic manipulation and thinking). However, close examination of existing theories reveals that the concept of covert mediating response always entails relative independence from environmental conditions. Since this last statement needs justification, the theories of Watson and Skinner will be briefly examined. Both men are identified in the minds of many psychologists with an empty organism, blackbox approach.

VERBAL MEDIATING RESPONSES AND SELF-STIMULATION

Watson and Skinner probably owe their reputation for rigidity about theoretical assumptions dealing with unobservables to unjustified claims to objectivity when talking about their theoretical systems. But it is not so much what they say *about* their theorizing as what they do in their theorizing that concerns us most in this discussion. Goss (1961) has argued convincingly that Watson's formulations of language learning and symbolic processes were forerunners of today's mediation models, and Watson's diagrams showing these relations in S–R terms "anticipate those which are now largely identified with Hull [p. 289]." Development of language habits, according to Watson, parallels the development of motor habits (manipulative, locomotive, and perhaps emotional, visceral responses). At first, motor habits alone are established. Later the individual finds himself in situations where both motor and labeling responses are called for. In

---

occurrence of the response is a sufficient and necessary condition for the occurrence of the self-stimulation.

the end, with internalization or elimination of other nonvocal responses, the stimuli may come to elicit vocal responses only.[3] These vocal responses may be either overt or covert. At this stage—and this is the crucial point for the present argument—the vocal responses may act as stimuli for (i.e., mediate) nonvocal muscular responses or other vocal responses. This property, by which these chained covert events (vocal and nonvocal) act as both stimuli and responses, is referred to here as *self-stimulation*, and it is this quality which renders the covert mediating responses relatively independent of environmental stimulation.

Watson's affirmation that these mediating responses are peripheral rather than central is of no crucial consequence to the present argument. The centrality-peripherality controversy which raged half a century ago has now subsided, not because the motor theory of thought has been disproved, but because neurologizing in psychology has become generally acceptable. Yet, even today, we find theorists, such as Osgood, unwilling to commit themselves to either position. The nature and locus of "the little r_m" remains unspecified (cf. Osgood, 1953, pp. 697–698; Osgood, Suci, & Tannenbaum, 1957, pp. 7–8).

Skinner's (1957) book, *Verbal Behavior*, is often cited as an example of the weakness of a simple S–R approach to language behavior and higher symbolic processes (e.g., Chomsky, 1959; Osgood, 1958). The criticisms leveled against Skinner's so-called single-stage S–R theory center around the notion that symbolic activity, meaning, semantic generalization, and the like, require inferred mediating processes that are relatively independent of overt stimulation. Despite what Skinner may say about the methodological difficulties of theorizing about unobservables and his stated reluctance to do so, the important point to realize is that he does make extensive use of inferred mediating processes. Again, this statement needs justification, and two examples will be cited here to support the argument. The first involves Skinner's interpretation of "meaning," while the second relates to his description and explanation of "semantic generalization."

MEANING AND SEMANTIC GENERALIZATION

The description of meaning which Skinner gives is remarkably similar to Osgood's sign conditioning paradigm—remarkable only because of the distance which is supposed to separate these two theorists. According to Osgood (1953, p. 696), the meaning of the word SPIDER is the representational mediation process (r_m) which becomes conditioned to the sign according to the classical Pavlovian conditioning paradigm:

$$\begin{tikzcd}
\text{spider-object} & \dot{S} \arrow[rrr] & & & R_T \arrow[dll, dash, dashed] & \text{unconditioned total reaction} \\
\text{the sign, SPIDER} & \boxed{S} \arrow[r] & r_m \arrow[r, dashed] & s_m \arrow[r] & R_X & \text{instrumental response independently conditioned}
\end{tikzcd}$$

In this diagram, we can distinguish Osgood's two stages. The first is the classical conditioning of a meaning response (r_m) to the verbal symbol SPIDER. The second, independent of the first, is an instrumentally conditioned response (R_X) to the self-stimulation (s_m). An example of R_X is the overt verbal response "snake".

Now compare to this Skinner's (1957) following account:

> But when we get someone to laugh by telling a funny story, we use stimuli conditioned according to the classical Pavlovian pattern. If a verbal stimulus frequently accompanies some state of affairs which is the unconditioned or previously conditioned stimulus for an emotional reaction, the verbal stimulus eventually evokes this reaction. Thus if one is afraid of snakes, and if the verbal stimulus

---

[3] Osgood (1961) has argued that the form of his representational mediation process (r_m) is nonverbal. In this respect, he differs from Watson or Bousfield, but not from Skinner, as will be shown below.

> "snake" has sometimes accompanied real snakes, the verbal stimulus alone may evoke an emotional reaction [p. 155].

Thus, verbal symbols may acquire the property of evoking emotional reactions which are representative of unconditioned responses not proper to the verbal stimulus as such. Furthermore, the overt verbal response, "snake", is independent of these emotional reactions, as attested by the following remarks:

> The emotional reaction is usually a by-product of some other verbal function. The verbal environment does not establish the response "snake" primarily to evoke such a reaction on the part of the listener [Skinner, 1957, p. 155].

In short, both Osgood's and Skinner's accounts involve the notion of two independent processes in meaning: classically conditioned representational responses to symbols, and instrumentally conditioned (operant) verbal responses. Perhaps because Skinner does not use diagrams, these relations are less clearly perceived in his, as compared to Osgood's, account.

With respect to the problem of semantic generalization, Skinner (1957, p. 155) cites Diven's (1937) classic demonstration of this phenomenon, while Osgood (1953, p. 701) describes in somewhat greater detail a number of such experiments involving generalization of the galvanic skin response. According to Osgood, generalization of an overt response from one symbol to another physically dissimilar symbol (e.g., STYLE to FASHION) is mediated by the common meaning response shared by the two symbols. Skinner's (1957) account of the same phenomenon, which he cites as an example of "metaphorical extension" is discussed in connection with the words that Shakespeare puts in Romeo's mouth: "Juliet is the sun." His argument runs as follows:

> Sometimes a genuine extension seems to occur when no similarity between stimuli expressible in the terms of physical science can be demonstrated. There are several possible explanations. Two stimuli may have a common effect upon the responding organism, which mediates the extension of the response. In the example "Juliet is the sun," it is possible that a physical similarity could not be plausibly established. Only to Romeo did Juliet glow with the light of the dawn. The metaphorical extension might have been mediated by, say, an emotional response which both the sun and Juliet evoked in him [p. 97].

The reader can judge for himself on the similarity of this account by Skinner and an account which Osgood would give of the same event.

THINKING

Skinner has been accused of shunning the problem of higher mental processes, and yet he has stated (p. 434) that an account of verbal behavior would not be complete until the problem of thinking is discussed. One should not be distracted by his warning that "verbal behavior which cannot be observed by others is not properly part of our field [p. 434]" when, in the same paragraph, he himself can disregard the warning as impractical:

> It is tempting to avoid the problems it raises by confining ourselves to observable events, letting anyone extend the analysis to his own covert behavior who wishes to do so. But there would then be certain embarrassing gaps in our account. In intraverbal chaining, for example, necessary links are sometimes missing from the observable data. When someone solves a problem in "mental arithmetic," the initial statement of the problem and the final overt answer can often be related only by inferring covert events. We also have to account for verbal behavior which is under the control of covert speech—which reports it (Chapter 5) or qualifies it with autoclitics (Chapter 12). Covert behavior has also had to be considered in discussing grammar (Chapter 13), sentence composition (Chapter 14), editing (Chapter 15), and other topics of Part V. Some discussion of its dimensions is therefore required [p. 434].

Skinner is not opposed to the logic of the centralist position: "Physiological processes mediate the probability of covert and overt responses alike . . . [p. 435]" even though he may not be inclined to make guesses about physiological mediators. He shares with Osgood, Maltzman, Staats, N. E. Miller, and others, the belief that "the data which give rise to the notion of covert speech can be dealt with the degree of rigor prevailing elsewhere in a science of verbal behavior at the present time [p. 435]."

It was stated above that the property of self-stimulation is a crucial characteristic of inferred mediating processes in that it allows for relative independence of environmental stimulation. According to Skinner:

> . . . the range of verbal behavior is roughly suggested, in descending order of energy, by shouting, loud talking, quiet talking, whispering, muttering "under one's breath," subaudible speech with detectable muscular action, subaudible speech of unclear dimensions, and perhaps even the "unconscious thinking" sometimes inferred in instances of problem solving [p. 438].

When it is realized that "subaudible speech of unclear dimensions" (neural?) may be "under the control of" or "mediated by" other elements of the same nature (i.e., unmeasurables), it then becomes quite clear that Skinner's account of thinking is as far removed from the external environment as any of the avowed mediationists. Self-stimulation is thus a central concept in Skinner's theory:

> The speaker's own verbal behavior automatically supplies stimuli for echoic, textual, or intraverbal behavior, and these in turn generate stimuli for further responses [p. 439].

> Since automatic reinforcement [provided by the speaker's own verbal behavior] need not respect the contingencies which prevail in the external environment, controlling relations can be "stretched" at will, beginning perhaps with a slight exaggeration but leading eventually to fiction and lying [pp. 441–442].

The importance of self-stimulation in Skinner's system is attested by the central role relegated to his concepts of "self-mand" (which functions as self-reinforcement), "self-tact" (which serves a self-controlling function), and "covert self-editing" (verbal behavior under the control of "private stimulation").

It may be argued (e.g., Chomsky, 1959) that the expression "under the control of private stimulation" represents a gratuitous usage of the term "control." This criticism, whether valid or not, is irrelevant to the present argument. Similar criticisms can be raised with respect to the circularity in the definition of S and R, the term, "reinforcement," the application of the concept of semantic generalization to the stylistic usage of metaphorical expressions, or to the use of the concept of response to refer to unspecified neural processes. Criticisms raised by Chomsky (1959) in his review of Skinner's *Verbal Behavior* represent a linguist's indictment of all types of S–R models in language behavior, Watsonian, Osgoodian, and Skinnerian alike. They will have to be dealt with elsewhere.[4] What is asserted here is that the Skinnerian concept of mediated self-stimulation has the same theoretical characteristics as Osgood's representational mediation process.[5]

The impression should not be given here that Skinner's critics are unaware of the inconsistency between his claim to objectivity and the use of the concept of mediated self-stimulation. Thus Osgood (1958), commenting on Skinner's use of the "extended tact" to explain semantic generalization states:

> To say that it occurs "covertly" cancels the pristine objectivity of the system and eliminates any real distinction between single-stage and two-stage models.

> It must be emphasized, however, that there is nothing necessarily incompatible between what Skinner has done here and an explication of representational (symbolic) processes. . . .

---

[4] The seriousness of Chomsky's remarks is indicated by the significant fact that 5 years after his influential indictment no major attempt on the part of S–R theorists has appeared in the psychological literature to answer his criticisms.

[5] This assertion should not be confused with a similar sounding conclusion recently formulated by the philosopher Fodor (1965), but which was based on an inadmissible reduction of Osgood's mediation theory to a fictional single-stage S–R paradigm.

> Having agreed that there are implicit nonverbal processes in behavior, Skinner does not go on to explicate their nature and function. This is the major insufficiency of *Verbal Behavior* . . . [p. 212].

Clearly, then, the charge is not one of incompatibility.

Skinner has "always regarded the last few chapters in *Verbal Behavior* as the most important" and believes that "the possibility of analyzing thinking is one of the major contributions of the book."[6] In view of this, it is ironical that his analysis of thinking (which appears as the last chapter in the book) has often been summarily dismissed as an afterthought inconsistent with his general system.

---

[6] B. F. Skinner, personal communication, February 1965.

REFERENCES

CHOMSKY, N. Review of *Verbal Behavior* by B. F. Skinner. *Language*, 1959, **35**, 26–58.

DIVEN, K. Certain determinants in the conditioning of anxiety reactions. *Journal of Psychology*, 1937, **3**, 291–308.

FODOR, J. A. Could meaning be an r_m? *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 73–81.

GOSS, A. E. Early behaviorism and verbal mediating responses. *American Psychologist*, 1961, **16**, 285–298.

OSGOOD, C. E. *Method and theory in experimental psychology*. Cambridge: Oxford University Press, 1953.

OSGOOD, C. E. A question of sufficiency. *Contemporary Psychology*, 1958, **3**, 209–212.

OSGOOD, C. E. Comments on Professor Bousfield's paper. In C. N. Cofer & B. Musgrave (Eds.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961. Pp. 91–106.

OSGOOD, C. E., SUCI, G., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

SKINNER, B. F. *Verbal behavior*. New York: Appleton-Century-Crofts, 1957.

(Received March 19, 1965)

Psychological Review
1966, Vol. 73, No. 4, 382–387

NEURAL CONSOLIDATION AND ELECTROCONVULSIVE SHOCK REEXAMINED [1]

JAMES L. McGAUGH AND LEWIS F. PETRINOVICH
University of California, Irvine
State University of New York, Stony Brook

In a recent paper Lewis and Maher (1965) concluded that the findings of electroconvulsive shock (ECS) studies provide little support for the consolidation hypothesis, and they proposed the hypothesis that ECS effects on behavior are due to conditioned inhibition. The present paper points out that the evidence cited by Lewis and Maher, contrary to their conclusion, provides strong support for the consolidation hypothesis and that the conditioned-inhibition hypothesis proposed by Lewis and Maher is inadequate on both logical and empirical grounds.

During the past 2 decades, a large number of studies have shown that electroconvulsive shock (ECS) impairs retention if the ECS treatment is administered within a short period of time after the original learning (Glickman, 1961; McGaugh, 1965). The most generally accepted interpretation of these findings is that the ECS treatment interferes with processes involved in the consolidation or storage of memory traces. In a recent review Lewis and Maher (1965) attempted to ". . . . take up the major experimental variables in which ECS has been involved and to evaluate the data derived from their manipulation insofar as it pertains to consolidation [p. 226]." They also posed the question: "Are there adequate data to substantiate the notion that a period of time following learning is necessary for an engram to be fixed [p. 226]?" These are, of course, two separate questions. The question of whether the findings of ECS have any implications for consolidation theory is one; the question of whether consolidation theory is supported by *any adequate data* is still another. In restricting their review to studies using ECS, Lewis and Maher could not, of course, attempt to answer the latter question. It is unfortunate that they did not consider their more general question, for there is a rather substantial amount of data indicating that memory storage is time dependent (e.g., McGaugh, 1965; McGaugh & Petrinovich, 1965). Most of this research has been published since the publication of Glickman's (1961) review of earlier evidence relevant to the consolidation hypothesis.

On the basis of their analysis of some of the ECS research literature, Lewis and Maher: (*a*) concluded that very little, if any, of the ECS literature supports the consolidation hypothesis, and (*b*) presented an alternative interpretation of ECS effects on behavior. In the present paper we point out that: (*a*) Much of the literature reviewed by Lewis and Maher is irrelevant to the consolidation hypothesis; (*b*) contrary to the conclusion of Lewis and Maher there is a substantial amount of quite adequate evidence that ECS affects memory storage; and (*c*) Lewis and Maher's alternative interpretation of the behavioral effects of ECS is inadequate on both empirical and logical grounds.

ANALYSIS OF LEWIS AND MAHER'S REVIEW OF ECS RESEARCH

The analyses presented in the introduction and review sections of the Lewis and Maher paper are, for the most part, irrelevant to the consolidation hypothesis.

---

[1] Supported by Research Grants MH-07015, MH-10261, and MH-06096-0351 from the National Institute of Mental Health, United States Public Health Service, and by Research Grant GB1612 from the National Science Foundation.

For example, Lewis and Maher (1965, p. 226) point out that much forgetting is caused by proactive inhibition and that this phenomenon is difficult to encompass within a consolidation framework. This is true. It is also irrelevant. It has not been claimed that interference with consolidation is the *sole* cause of "forgetting" (i.e., failure to recall); it has only been claimed that it may be *one* cause of forgetting. Consolidation theory is not only a hypothesis of failure to recall, it is also (and primarily) a hypothesis of memory storage. And, as such, it has not yet been completely abandoned even by researchers in verbal learning (e.g., Madsen, 1963; Walker, 1963).

In the last section of their review, Lewis and Maher acknowledged that several experimenters (e.g., Hudspeth, McGaugh, & Thomson, 1964 [2]) have pointed out that the behavioral effects of a series of ECS treatments are quite different from those produced by a single ECS treatment. Retrograde amnesia can be produced with a single ECS treatment. Other effects (including aversive effects) are produced when ECS is administered repeatedly. In view of the fact that repeated ECS treatments are known to produce both amnesia and other effects which influence behavior, such studies cannot provide the best test of the consolidation hypothesis. Unfortunately most of the experiments cited by Lewis and Maher involved the use of multiple ECS treatments. It is clear that unconfounded evidence concerning memory consolidation must be obtained from experiments in which only a single ECS treatment is administered (or at least in which the effects of the first and subsequent treatments can be readily distinguished). Since this problem was first clearly recognized, a large number of relevant studies have been published. In addition to those cited by Lewis and Maher (Chorover & Schiller, 1965; Heriot & Coleman, 1962; Hudspeth et al., 1964; Pearlman, Sharpless, & Jarvik, 1961; Weissman, 1963), there are others (Chevalier, 1965; King, 1965; Leonard & Zavala, 1964; Madsen & McGaugh, 1961; McGaugh & Madsen, 1964; Quartermain, Paolino, & Miller, 1965; Tenen, 1965; Weissman, 1964). Since all of these studies found clear evidence that retrograde amnesia is produced by a single ECS, Lewis and Maher's (1965) conclusion, "It seems quite possible that the consolidation notion has not yet been tested at all [p. 237]," is obviously unjustified.

Lewis and Maher further state that studies using more than one training trial are irrelevant to the consolidation hypothesis. We do not agree with this conclusion. On the contrary, the series of experiments by Thompson and his colleagues (e.g., Thompson & Dean, 1955; Thompson & Pennington, 1957) has shown that a single CS produces retrograde amnesia if the treatment is administered within a short period after animals are trained (with massed trials) to a relatively stringent criterion. And, as would be expected according to the consolidation hypothesis, the amount of amnesia produced by a posttraining ECS treatment varied inversely with the length of the intertrial interval used during the training (Thompson & Pennington, 1957).

In discussing this latter experiment, Lewis and Maher argue that the effect of massed acquisition trials is due to temporary performance variables rather than to ones involved in learning. Contrary to the conclusion of Lewis and Maher, the evidence that the effect of massed practice is but a temporary one due to performance variables is not at all strong. If anything, distribution of practice generally results in superior retention as well as superior learning (Deese, 1958; Madsen, 1963).

On the basis of findings of numerous studies using a single ECS treatment, it is known that the length of the posttraining interval during which an ECS treatment will produce some degree of amnesia varies considerably as a function of many variables, including task complexity, degree of training, and age (Glickman,

[2] Incorrectly cited by Lewis and Maher (1965, p. 233) as McGaugh and Hudspeth, but not cited in the references of their paper.

1961). Lewis and Maher pointed out that Chorover and Schiller (1965) recently reported an experiment in which ECS produced retrograde amnesia only when administered within ½ minute after a training trial. The fact that memory storage with this task and conditions was somewhat more rapid than that obtained in other research is interesting, but it does not mean, as Lewis and Maher suggest, that this particular study ". . . is the only ECS study relevant to consolidation [p. 233]." Relevancy of experiments is determined by examining the appropriateness of the designs of experiments, not their outcomes. And, as we have indicated, numerous appropriately designed (i.e., relevant) studies have provided extremely strong evidence that memory storage is time dependent.

At the conclusion of the review section Lewis and Maher wrote:

> It seems that the pressure of the ECS data . . . is increasingly restricting the arena in which consolidation may be tested. It is not inconceivable that the possibility of such tests may soon vanish [p. 233].

Thus they seem to be doubting that there are any data requiring an explanation. In view of this, it is puzzling why, in the next section, they chose to present a new hypothesis concerning the basis of the amnesic effects of ECS.

CONDITIONED-INHIBITION HYPOTHESIS

The hypothesis proposed by Lewis and Maher is essentially as follows. ECS is assumed to be an unconditioned stimulus which produces an unconditioned response consisting of loss of consciousness. A weak version of the loss of consciousness, that is, relaxation, becomes conditioned to the cues present at the time of ECS administration. Apparent amnesia results, according to this view, because the relaxation response (or conditioned inhibition) competes with other responses.[3]

Any new hypothesis or theory has two functions to perform. First, it should provide an explanation of existing data, and, if possible, it should resolve some seemingly conflicting observations. At the very least, it should not ignore available data. Beyond that, any new hypothesis should have a heuristic role: It should suggest new experiments or new observations. Unfortunately Lewis and Maher made little effort to show in any detail how the conditioned-inhibition hypothesis might explain the diverse findings available at the time the paper was prepared. Beyond that, the hypothesis fails completely to explain at least one significant finding reported since the paper was originally prepared. On both logical and empirical grounds the hypothesis appears to be unable to provide explanations of a variety of findings.

1. The conditioned-inhibition hypothesis offers no explanation of the *differential* effects of single and repeated ECS treatments. There is extremely strong evidence that retrograde amnesia is produced by a single ECS treatment administered shortly after learning. *Other* effects (e.g., aversive effects and selective inhibition of a conditioned emotional response) are only found when ECS is administered repeatedly (e.g., Brady & Hunt, 1955; Hudspeth et al., 1964; McGaugh & Madsen, 1964). It is not at all clear how the conditioned-inhibition hypothesis would explain these findings.

2. The conditioned-inhibition hypothesis assumes that the inhibition is "conditioned to the environmental stimuli present at the time the coma was induced [p. 236]." The fact is that in most ECS experiments the ECS is administered *outside* of the training apparatus. Further, in recent direct tests of this hypothesis (Leonard & Zavala, 1964; Quartermain et al., 1965) the amount of retrograde amnesia produced by a single ECS has been found to be unaffected by variations in the place in which the treatment is administered. This evidence is *very* difficult to explain in terms of the conditioned-inhibition hypothesis.

3. The conditioned-inhibition hypothesis

---

[3] In an earlier version of this general hypothesis (Lewis & Adams, 1963) the amnesia was assumed to be due to conditioning of a partial form of a convulsion to cues present during ECS administration.

relies solely upon evidence from studies in which amnesia is indexed by *increases* in latencies. Failure to move or failure to move within a certain period of time is taken as evidence of amnesia in tasks requiring the animal to move in order to avoid a punishment or receive a reward. Thus, the findings of Duncan (1949) may be interpreted in this way. However, in most (but not all) recent studies using ECS the animals were trained to *inhibit* a response in order to avoid a punishment. According to the conditioned-inhibition hypothesis, ECS should enhance inhibitory avoidance learning (so-called passive avoidance); that is, the latencies should *increase* following an ECS treatment. A single ECS administered in the experimental apparatus does not produce increased latencies. However, a single posttrial ECS treatment even when administered outside of the apparatus clearly impairs inhibitory avoidance learning (e.g., Heriot & Coleman, 1962; Hudspeth et al., 1964; King, 1965; Madsen & McGaugh, 1961). In addition, there is extensive evidence (e.g., Glickman, 1961) that ECS produces amnesia in discrimination-learning tasks where differential responding rather than latency is used to measure learning and memory. Any hypothesis concerning the amnesic affects of ECS must consider these findings.

4. According to the conditioned-inhibition hypothesis, posttrial ECS treatments should, by inducing relaxation, facilitate extinction. However, ECS treatments have been found to impair extinction of both approach and avoidance responses (Gerbrant & Thomson, 1964; Madsen & Luttges, 1963).

5. Finally, Lewis and Maher (1965) concluded that: "Any procedure such as anesthesia that renders the animal unconscious before passage of current will prevent the establishment of an ECS-produced inhibition . . . [p. 237]." Amnesia would thus not be expected with such an experimental procedure. As Lewis and Maher pointed out, anesthetics can be used to prevent the occurrence of convulsions produced by electroshock stimulation, and the evidence does indicate that the effects produced by a *series* of ECS treatments are not found when the electroshock stimulation is administered while the animals are anesthetized (e.g., Friedman, 1953). However, the results of recent experiments clearly indicate that the retrograde amnesia produced by a *single* ECS is obtained even when the electroshock stimulation is delivered while the animals are anesthetized. The results of one study (McGaugh & Alpern, 1966) are shown in Figure 1. The left side of Figure 1 shows the performance on a second trial of mice given a foot shock on a previous day as they stepped from a small platform through a small hole into a darkened box. The performance of mice given an ECS while lightly anesthetized with ether (resulting in no convulsion) was comparable to that of animals given a normal ECS treatment. Both groups showed less retention than groups given the same treatments 1 hour after the trial. The amnesia found with ECS appears to be due to the electrical current rather than to the convulsions. Control groups shown on the right side of Figure 1 indicate that when no foot shock is given on the first trial, the treatments of ether, ECS, and ECS under anesthesia do not affect retention. Comparable results with rats and with other anticonvulsants have been obtained by Weissman (1965). Thus, this evidence indicates quite clearly

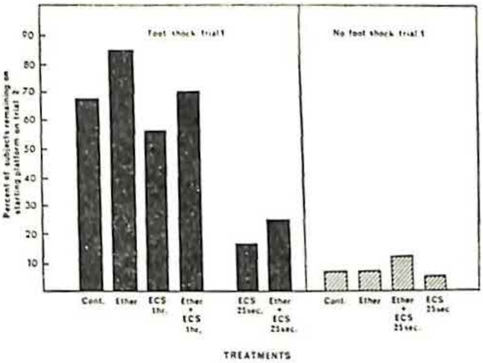

FIG. 1. Effects of posttrial treatments of ECS, ether anesthesia, and ECS delivered during ether anesthesia on one-trial avoidance learning in mice. (From McGaugh & Alpern, 1966. See text for details.)

that anesthesia *does not,* as Lewis and Maher conclude, prevent retrograde amnesia. Anesthesia does, however, seem to attenuate the pseudoamnesia as well as the aversive effects produced by multiple ECS treatments (Friedman, 1953).

CONCLUSION

We agree with Lewis and Maher's (1965) observation that: "The effects of ECS are undoubtedly very complicated [p. 237]." However, it seems highly unlikely that the conditioned-inhibition hypothesis produced by Lewis and Maher will be shown to be capable of explaining the diverse effects of ECS on behavior. In particular the conditioned-inhibition hypothesis seems unable to account for the *details* of studies of the amnesic effects of ECS. The conditioned-inhibition hypothesis, although perhaps superficially appealing, is clearly inadequate on both logical and empirical grounds. Studies of the effects of ECS on learning continue to provide support for the hypothesis that memory storage is time dependent. Thus, the consolidation hypothesis not only ". . . remains appealing" (p. 237), it remains supported. It does not appear that the conditioned-inhibition hypothesis will fare as well.

REFERENCES

BRADY, J. V., & HUNT, H. F. An experimental analysis of emotional behavior. *Journal of Psychology,* 1955, **40,** 313–324.

CHEVALIER, J. A. Permanence of amnesia after a single posttrial electroconvulsive seizure. *Journal of Comparative and Physiological Psychology,* 1965, **59,** 125–127.

CHOROVER, S. L., & SCHILLER, P. H. Short-term retrograde amnesia in rats. *Journal of Comparative and Physiological Psychology,* 1965, **59,** 73–78.

DEESE, J. *The psychology of learning.* New York: McGraw-Hill, 1958.

DUNCAN, C. P. The retroactive effects of electroshock on learning. *Journal of Comparative and Physiological Psychology,* 1949, **42,** 32–44.

FRIEDMAN, M. H. Electroconvulsive shock as a traumatic (fear-producing) experience in the albino rat. *Journal of Comparative and Physiological Psychology,* 1953, **48,** 555–562.

GERBRANT, L. K., & THOMSON, C. W. Competing response and amnesic effects of electroconvulsive shock under extinction and incentive shifts. *Journal of Comparative and Physiological Psychology,* 1964, **58,** 208–211.

GLICKMAN, S. E. Perseverative neural processes and consolidation of the neural trace. *Psychological Bulletin,* 1961, **58,** 218–233.

HERIOT, J. T., & COLEMAN, P. D. The effects of electroconvulsive shock on retention of a modified "one-trial" conditioned avoidance. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 1082–1084.

HUDSPETH, W. J., MCGAUGH, J. L., & THOMSON, C. W. Aversive and amnesic effects of electroconvulsive shock. *Journal of Comparative and Physiological Psychology,* 1964, **57,** 61–64.

KING, R. A. Consolidation of the neural trace in memory: Investigation with one-trial avoidance conditioning and ECS. *Journal of Comparative and Physiological Psychology,* 1965, **59,** 283–284.

LEONARD, D. J., & ZAVALA, A. Electroconvulsive shock: Retrograde amnesia and the single shock method. *Science,* 1964, **146,** 1073–1074.

LEWIS, D. J., & ADAMS, H. E. Retrograde amnesia from conditioned competing responses. *Science,* 1963, **141,** 516–517.

LEWIS, D. J., & MAHER, B. A. Neural consolidation and electroconvulsive shock. *Psychological Review,* 1965, **72,** 225–239.

MADSEN, M. C. Distribution of practice and level of intelligence. *Psychological Reports,* 1963, **13,** 39–42.

MADSEN, M. C., & LUTTGES, M. W. Effect of electroconvulsive shock on extinction of an approach response. *Psychological Reports,* 1963, **13,** 225–226.

MADSEN, M. C., & MCGAUGH, J. L. The effect of ECS on one-trial avoidance learning. *Journal of Comparative and Physiological Psychology,* 1961, **54,** 522–523.

MCGAUGH, J. L. Facilitation and impairment of memory storage processes. In D. P. Kimble (Ed.), *The anatomy of memory.* Palo Alto, Calif.: Science & Behavior Books, 1965. Pp. 240–291.

MCGAUGH, J. L., & ALPERN, H. Effects of electroshock on memory: Amnesia without convulsions. *Science,* 1966, **152,** 665–666.

MCGAUGH, J. L., & MADSEN, M. C. Amnesic and punishing effects of electroconvulsive shock. *Science,* 1964, **144,** 182–183.

MCGAUGH, J. L., & PETRINOVICH, L. F. Effects of drugs on learning and memory.

International Review of Neurobiology, 1965, **8**, 139–196.

PEARLMAN, C. A., JR., SHARPLESS, S. K., & JARVIK, M. E. Retrograde amnesia produced by anesthetic and convulsant agents. *Journal of Comparative and Physiological Psychology*, 1961, **54**, 109–112.

QUARTERMAIN, D., PAOLINO, R. M., & MILLER, N. E. A brief temporal gradient of retrograde amnesia independent of situational change. *Science*, 1965, **149**, 1116–1118.

TENEN, S. S. Retrograde amnesia from electroconvulsive shock in a one-trial appetitive learning task. *Science*, 1965, **148**, 1248–1250.

THOMPSON, R., & DEAN, W. A further study on the retroactive effects of ECS. *Journal of Comparative and Physiological Psychology*, 1955, **48**, 488–491.

THOMPSON, R., & PENNINGTON, D. F. Memory decrement produced by ECS as a function of the distribution of original learning. *Journal of Comparative and Physiological Psychology*, 1957, **40**, 401–404.

WALKER, E. L. Memory storage as a function of arousal and time. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 113–119.

WEISSMAN, A. Effect of electroconvulsive shock intensity and seizure pattern on retrograde amnesia in rats. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 806–810.

WEISSMAN, A. Retrograde amnesia effects of supramaximal electroconvulsive shock on one-trial acquisition in rats: A replication. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 248–250.

WEISSMAN, A. Effect of anticonvulsant drugs on electroconvulsive shock-induced retrograde amnesia. *Archives internationales de pharmacodynamic et de thérapie*, 1965, **154**, 122–130.

(Received July 22, 1965)

Psychological Review
1966, Vol. 73, No. 4, 388–392

ELECTROCONVULSIVE SHOCK AND INHIBITION:

SOME PROBLEMS CONSIDERED[1]

DONALD J. LEWIS
Rutgers University

AND

BRENDAN A. MAHER
University of Wisconsin

Retrograde amnesia due to electroconvulsive shock (ECS) is a memory or response deficit for events that have occurred somewhat prior to the convulsion. While it is possible that the amnesia is due to the failure of an engram to consolidate, in a recent paper Lewis and Maher (1965) pointed out that multiple ECS studies, which have so frequently been taken as supportive of the consolidation point of view, are not relevant. Here some criticisms of that paper are answered, and the weaknesses of many of the single ECS studies for this hypothesis are pointed out. A conditioned-inhibition hypothesis for the amnesic effect of ECS is also defended.

Some of the questions that have been raised by McGaugh and Petrinovich (1966) require careful answer. Others may be disposed of rather briefly. We shall consider the latter group first and save the main bulk of the space at our disposal to examine the central issues.

First, we must draw attention to the fact that by separating two sentences that appeared together as a continuous proposition in the original paper (Lewis & Maher, 1965) our critics have managed to convince themselves that we fail to appreciate that the consolidation hypothesis might draw support from studies unrelated to the electroconvulsive shock (ECS) technique. At the risk of tedious repetition, we would like to repeat the original statement as it was written:

> Our approach will be to take up the major experimental variables in which ECS has been involved and to evaluate the data derived from their manipulation insofar as it pertains to consolidation. Are there adequate data to substantiate the notion that a period of time following learning is necessary for an engram to become fixed [p. 226]?

These sentences, considered separately, might have referred to two separate questions. In the context and sequence in which we wrote them, it is quite clear that they refer to one question—namely, the relationship of ECS data to the problem of the neural consolidation hypothesis.

The second matter is their disagreement with our statement that the evidence is quite strong that the effect of massed trials is due to performance rather than learning variables. Legitimate differences of opinion are possible, but not frequent, on this topic, and we believe the better evidence and opinion support our point of view. In discussing this matter, Kimble (1961, p. 125) says:

> The evidence that what is involved is a performance factor comes from a variety of demonstrations that the effect of massed practice is temporary and that the decremental process dissipates spontaneously with rest.

In any case, faith in the effects of trial distribution is a shaky peg on which to hang the consolidation point of view.

Third, McGaugh and Petrinovich raise a question concerning the adequacy of our coverage of the ECS literature. It is generally difficult for writers to cite papers that have not yet appeared in print. For McGaugh and Petrinovich to state that we did not consider the studies of Chevailier (1965), King (1965), Quartermain, Paolino, and Miller (1965), and Tenen (1965) is quite accurate. For them to complain about it and to add that

---

[1] This paper was supported in part by Grant MH-07129 from the National Institute of Mental Health.

therefore our conclusions were "obviously unjustified" is simply silly.

McGaugh and Petrinovich want us to believe that the existence of retrograde amnesia itself must be taken as supporting a consolidation notion. We must make it clear that the question at issue is not whether ECS produces retrograde amnesia—if by this is meant impairment in performance of a previously learned response. The data are generally very clear on this point: ECS does tend to lower—at times to zero—frequencies of responses that have been acquired at some brief interval before the convulsion. The question is not the reality of the behavioral deficit; it is the question of an adequate theoretical account of its determinants. Thus the claim that "all these studies found clear evidence that retrograde amnesia is produced by a single ECS [p. 383]" is misleading. These studies do demonstrate a performance impairment, but they are not conclusive to the more precise issue of whether the performance impairment (retrograde amnesia) is a consequence of disruption of engram consolidation, or whether it is due to interference by some competing response.

McGaugh and Petrinovich refer to the more recent studies using only one ECS as if the conclusions they present are somehow unequivocally in favor of the consolidation hypothesis. This is by no means the case. Tenen (1965) studied the effects of single ECS trials on rats that had previously approached a waterhole while thirsty and received reinforcement there. Two ECS intervals were used, 12 seconds and 3 hours. Examination of the rate of responding showed that ECS, predictably, was followed by a diminution of responding compared to groups that did not receive ECS. However, the differences between the 12-second and 3-hour groups were not significant—a finding that prompted Tenen to suggest that "it seems possible that ECS might exert some limited retrograde amnesic effects even three hours after reinforcement [p. 2]." But it also appears from examination of graphic data presented in this study that the amnesia found in the delayed group was not different from that exhibited by a group of rats given simple painful foot shock in lieu of ECS. Chorover and Schiller (1965a) also criticized the consolidation interpretation given to the delayed group. The data hardly constitute univocal support for an engram consolidation hypothesis.

A very large proportion of the experiments using one ECS employ a situation in which a rat is placed on a raised platform. When he steps off, he receives a foot shock that causes him to remain on the platform when returned there for the next trial. If ECS follows the foot shock, however, the rat seems to forget the punishment and steps off the platform directly on the next trial. Thus, memory of punishment has apparently been disrupted, and interference with consolidation is the typical guess as to the mechanism for the disruption. But this conclusion is far from demanded by the data. Routtenberg and Koy (1965) showed that one ECS will increase the tendency of a rat to step off a platform independent of any avoidance training associated with it and specifically called for a reevaluation of the previous literature (p. 288). Chorover and Schiller (1966) showed that a single ECS, as administered in typical consolidation studies, will produce a relative increase in locomotor activity. This increase in activity results in a marked reduction in avoidance and leads to the frequent interpretation that an engram has failed to consolidate. The effect of ECS in these situations is indeed time dependent; the question is whether consolidation is involved. We tend to think that the evidence does not warrant the conclusion that it is.

INHIBITION HYPOTHESIS

Let us now turn to the conditioned-inhibition hypothesis. It is not easy to see why McGaugh and Petrinovich find this hypothesis even "superficially attractive," as it is quite apparent that they do not understand it. The purpose of the hypothesis is to attempt the explanation of *ECS data*, whether or not the ECS studies have been conducted in a manner germane to the consolidation theory of memory. Our critics accept that multiple ECS studies are not very helpful in testing consolidation notions. We agree. However, this does not mean that the consequences of multiple ECS treatment are not of psychological interest themselves. The amnesic phenomena produced by multiple ECS are just as "real" as those produced by one ECS, but they cannot readily be accounted for in terms of engram destruction as many have recently believed (Thomson, McGaugh, Smith, Hudspeth, & Westbrook, 1961).

McGaugh and Petrinovich are worried by the failure of the inhibition hypothesis to explain the differential effects of single and repeated ECS treatments. They believe that effects such as fear are found only when ECS is administered repeatedly. Reference to King (1965), cited by them, shows that such effects are in fact obtained by a single ECS treatment, although King also states that fear produced by a single ECS cannot be the sole explanation for the amnesia.

The Quartermain et al. (1965) study is also cited by McGaugh and Petrinovich as negative to our interpretation. In our paper we said that there was only one study (Chorover & Schiller, 1965b) in the literature with a learning-ECS interval sufficiently short to test the consolidation notion. Now we will amend that statement to include the Quartermain et al. (1965) report. *If* the consolidation interpretation has any support from the ECS literature, it comes from these two studies, and our objections to the others still stand.

Another criticism is perhaps more baffling. McGaugh and Petrinovich state that: "the conditioned-inhibition hypothesis relies solely upon evidence from studies in which amnesia is indexed by *increases* in latencies [p. 384]." This could mean that the studies mentioned in our original paper (Lewis & Maher, 1965) under the specific heading, "A Conditioned Inhibition Hypothesis," include only such studies. A careful scrutiny of pages 234–237 of our paper shows this not to be the case. Alternatively, perhaps McGaugh and Petrinovich mean that the whole paper rests only on such studies. This, clearly, is not the case either as the citations to the work of Brady and his colleagues point specifically to the use of increases in rate of bar press as measures of ECS effects. The study by Williams (1961) points to the effects of ECS in reducing hesitations in a running response. Many other kinds of response are included in our review. As the criticism is somewhat unclear, we might venture to suggest what the problem is. McGaugh and Petrinovich incorrectly deduce that conditioned inhibition should enhance inhibitory avoidance learning. Inhibitory avoidance learning —of which freezing in the conditioned emotional response pattern is the exemplar—involves the maintenance of considerable muscular tension in a highly aroused animal. Under conditions where the fear is gradually reduced—as by normal extinction processes—the reduction in arousal is evidenced by an *increase* in the general locomotor activity. Thus, the animal that is in a lowered state of arousal should be less likely to "freeze"—more likely to move around—than one that is in a highly fearful state. The effect of the conditioned inhibition, we hypothesize, is to *lower arousal levels*, not to increase those motor inhibitory patterns that are invoked in passive avoidance situations. In a word, the difference is between the implications of *protective inhibition* (a state on the general arousal-inhibitory continuum) and the inhibition of specific motor responses, a matter that depends, of course, upon the pattern of muscular innervation that a given stimulus may excite. It is a pity that this confusion could not have been avoided by our critics. We had anticipated that the clear reference to *protective* inhibition (Lewis & Maher, p. 234) would have obviated this problem. However, it may help at this point to reiterate that under some conditions the shift from high arousal to lower states may be accomplished by an increase in locomotor activity where the locomotor activity was previously being inhibited as part of the high-arousal pattern. Thus the hypothesis does not depend upon increased latencies only; it may also be used to account for the reduction in the latency of a response which had been suppressed by punishment.

Contrary to the simplified deductions of McGaugh and Petrinovich, the conditioned-inhibition hypothesis does not inevitably compel the prediction that posttrial ECS treatments will facilitate extinction. Predictions cannot be made in the absence of specific information about the levels of arousal appropriate to the production of the response and the processes involved in extinction. For example, to the extent that extinction depends upon the occurrence of the response, this being followed by nonreinforcement, then anything that depresses the rate of response may serve to slow the occurrence of extinction trials, their spacing, etc.

Finally, we should note that the brief account of an experiment using ether anesthesia permits us to do no more than raise some questions. One crucial question relates to the interval between the occurrence of the foot shock and the administration of ether. From the description we must assume that the animals in the Ether-ECS/25 seconds condition were anesthetized at some posttrial interval of less than 25 seconds. Thus the unconscious state was presumably induced fairly rapidly after foot shock. The finding that this produces retrograde amnesia is in line with the work of Pearlman, Sharpless, and Jarvik (1961), cited in the original review. As we have already pointed out (Lewis & Maher, 1965, p. 236), ". . . if the unconsciousness induced by the anesthetic is rapid, and done in the experimental situation, then it should have the same retrograde amnesic effects that ECS alone does." We have not advanced the view that the *convulsion* is necessary to produce the conditioned-inhibition paradigm—simply that rapid induction of unconsciousness is necessary.

We do have some points of agreement with McGaugh and Petrinovich. We agree that memory (and hence learning) must depend upon physiological storage processes, and that these will have some time-bound characteristics. We agree that multiple ECS treatments are a poor way to test this notion. We agree that there are many determinants of forgetting other than simple response competition, just as there are many ways to account for performance impairment besides the assumption that an engram has been destroyed.

We disagree with them on many other counts. They are unhappy, for example, with conditioned inhibition as an explanation of ECS effects. In the case of multiple ECS treatments they are no longer offering the consolidation hypothesis as an alternate (e.g., see McGaugh, Westbrook, & Thomson, 1962, who discuss multiple ECS studies as supporting "perseveration-consolidation"). They are not, in fact, offering any alternative. Until they, or others, do we must point out that the inhibition hypothesis does explain an impressive bulk of experimental data and resolves many of the contradictions that arose when these experiments were interpreted in engram terms.

McGaugh and Petrinovich feel that we have been unaware of the results of single ECS studies. Examining those studies they have cited, we find that they do report performance impairment; they do not provide experimentally crucial evidence for the consolidation hypothesis. The studies are generally compatible with the conditioned-inhibition hypothesis.

Under the circumstances we shall continue to explore the conditioned-inhibition hypothesis with every expectation that empirical data will provide the corrective amendments that hasty criticism does not.

CONCLUSION

In our previous article we reviewed the ECS literature which, for about 20 years, had been taken, with very few exceptions, as confirming the notion of neural consolidation. It was argued that the literature was largely irrelevant to the issue and that only the very few experiments using one ECS could be considered as relevant. Among these experiments, only one (Chorover & Schiller, 1965b) had worked with a sufficiently short interval between learning and convulsion to be clearly within a possible consolidation period. This one study had found no behavioral impairment beyond a 10-second interval. This interval has since been confirmed by a second experiment, and serious doubt remains that the impairment found in other experiments using one ECS is due to interference with a consolidating engram. In the present paper we have defended our point of view from some hasty criticism, pointed out some of the reasons why most single ECS studies are suspect, and defended our hypothesis of conditioned inhibition for the retrograde amnesic effects of ECS.

REFERENCES

CHEVALIER, J. A. Permanence of amnesia after a single posttrial electroconvulsive seizure. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 125–127.

CHOROVER, S. L., & SCHILLER, P. H. Retrograde amnesia. *Science,* 1965, **129**, 1521. (a)

CHOROVER, S. L., & SCHILLER, P. H. Short-term retrograde amnesia in rats. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 73–78. (b)

CHOROVER, S. L., & SCHILLER, P. H. Reexamination of prolonged retrograde amnesia in one-trial learning. *Journal of Comparative and Physiological Psychology,* 1966, **61**, 34–41.

KIMBLE, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

KING, R. A. Consolidation of the neural trace in memory: Investigation with one-trial avoidance conditioning and ECS. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 283–284.

LEWIS, D. J., & MAHER, B. A. Neural consolidation and electroconvulsive shock. *Psychological Review,* 1965, **72**, 225–239.

McGAUGH, J. L., & PETRINOVICH, L. F. Neural consolidation and electroconvulsive shock reexamined. *Psychological Review,* 1966, **73**, 382–387.

McGAUGH, J. L., WESTBROOK, W. H., & THOMSON, C. W. Facilitation of maze learning with posttrial injections of 5-7-diphenyl-1-3-diazadamanton-6-07 (1757 I.S.). *Journal of Comparative and Physiological Psychology,* 1962, **55**, 710–713.

PEARLMAN, C. A., JR., SHARPLESS, S. K., & JARVIK, M. E. Retrograde amnesia produced by anesthetic and convulsant agents. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 109–112.

QUARTERMAIN, D., PAOLINO, R. M., & MILLER, N. E. A brief temporal gradient of retrograde amnesia independent of situational change. *Science,* 1965, **149**, 1116–1118.

ROUTTENBERG, A., & KOY, K. E. Effect of one electroconvulsive seizure on rat behavior. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 285–288.

TENEN, S. S. Retrograde amnesia from electroconvulsive shock in a one-trial appetitive learning task. *Science,* 1965, **148**, 1248–1250.

THOMSON, C. W., McGAUGH, J. L., SMITH, C. E., HUDSPETH, W. J., & WESTBROOK, W. H. Strain differences in the retrograde effects of electroconvulsive shock on maze learning. *Canadian Journal of Psychology,* 1961, **15**, 69–74.

WILLIAMS, G. J. The effect of electroconvulsive shock on an instrumental conditioned emotional response ("conflict"). *Journal of Comparative and Physiological Psychology,* 1961, **54**, 633–637.

(Received December 20, 1965)

VOL. 73, NO. 5

SEPTEMBER 1966

PSYCHOLOGICAL REVIEW

A MARKOV MODEL FOR CLASSICAL CONDITIONING: APPLICATION TO EYE-BLINK CONDITIONING IN RABBITS[1]

JOHN THEIOS AND JOHN W. BRELSFORD, JR.[2]

University of Texas

A 3-state Markov model is postulated for classical conditioning. Specific implications of the model for eye-blink conditioning are discussed. The Markov model, a general response-independent model, and a single-operator linear model with individual differences were applied to eye-blink conditioning data of rabbits. Likelihood ratio analyses of individual response sequences indicate that the Markov model is far superior to the other models as a description of the conditioning process in rabbits. Speculations concerning theoretical identifications of the abstract states of the Markov process are made. Application of the Markov model to human eyelid conditioning is discussed.

Noxious stimulation of the eye of the rabbit results in a reflex response which is characterized by retraction of the eyeball into its orbit, extension of the nictitating membrane partially across the cornea, and closure of the outer eyelids. The work of Gormezano and his associates has shown that each of these three components of the rabbit eye-blink reflex may be reliably conditioned to a neutral stimulus and that the three response components are very highly correlated (Deaux & Gormezano, 1963; Gormezano, Schneiderman, Deaux, & Fuentes, 1962; Schneiderman, Fuentes, & Gormezano, 1962). Their research indicates that paired presentations of a conditioned stimulus (CS), for example, a tone, and an unconditioned stimulus (UCS), for example, an air puff to the eye, result in an increase in the probability of an eye blink to the CS. However, presentations of the CS alone, UCS alone, or unpaired presentations of CS and UCS, result in no systematic changes in the probability of a blink during the recording period.

The purpose of this research was to develop a theoretical description of the changes which take place in the trial-by-trial probability of response in individual rabbits during an eye-blink conditioning experiment. Extension of the nictitating membrane was classically conditioned in 101 rabbits. The method and procedure is described in Brelsford and Theios (1965), where data from 70 of the 101 rabbits are presented. The subjects

---

[1] This research was supported by PHS Research Grants HD 00950-01, 02, 03 from the National Institute of Child Health and Human Development, Public Health Service. We wish to express our appreciation to Rodney W. McGinnis for developing the computer programs for obtaining the likelihood values reported in this paper.

[2] Now at Stanford University.

(*S*s) were given 150 trials of conditioning in a single session with a .25-second CS-UCS interval and a constant 111-second intertrial interval. The CS was an 88-decibel buzzer, and the UCS was a single 1-millisecond spike of electric shock, delivered across the rabbit's right eye. The external eyelids of the rabbit were not clamped open. Thus, the response involved extension of the nictitating membrane, closure of the outer eyelids, and retraction of the eyeball.

Under these conditions, the initial probability of a response to the CS in the .25-second observation interval after CS onset is effectively zero. After conditioning, however, rabbits have an asymptotic probability of about .95 of blinking to the CS in the .25 second interval before the UCS onset. Operationally, a response to the CS which occurs during the observation interval will be designated a conditioned response (CR). Just as in the experiments by Gormezano and his associates, the group mean probability of a CR increases as a slightly **S**-shaped function of the number of paired CS-UCS trials.

RESPONSE-INDEPENDENT MODELS

In situations such as classical conditioning, in which the stimulus events are presented on every trial, there are two major approaches which may be taken with respect to the construction of learning models. The first approach is to assume that the theoretical response probabilities are described by the ordered set $\{p_1, p_2, \ldots p_n, \ldots p_M\}$, where p_n represents the probability of a CR on any Trial n, for $n = 1, 2, \ldots, M$. With this type of model, the response given on Trial n depends only upon the value of p_n and is independent of the sequence of responses given on trials prior to n. Models having this independence property include the single-operator linear model of Bush and Mosteller (1955) and Estes (1950), the single-operator beta model of Luce (1959), and the response-strength model of Spence (1956). In the first stages of model construction it is often convenient to assume that the *S*s are homogeneous with respect to the values of the parameters of the model. Assuming homogeneity of parameters, the single-operator linear model predicts that the probability of a CR for individual *S*s as well as for the group should be an exponentially increasing function of trials. Since the obtained group mean functions for rabbit eye-blink conditioning are typically **S** shaped (Schneiderman et al., 1962), the assumption of homogeneity of parameters cannot be made if one is to accurately predict the mean learning curve by the single-operator linear model. The homogeneity assumption can be made with respect to the strength models of Luce (1959) and Spence (1956) since these models permit **S**-shaped conditioning curves for individual *S*s.

MARKOV MODEL

As an alternative to the response-independent models, it may be postulated that response probability on Trial n is a function of the sequence of responses which occurred on the previous trials. Because the class of models having this dependence property is so large, our consideration will be restricted to the specific response-dependent Markov model defined by the following transition matrix, response vector, and starting vector:

$$\begin{array}{c} \\ \begin{matrix} C \\ A \\ N \end{matrix} \end{array} \begin{array}{c} \begin{matrix} C & A & N \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ c & 1-c & 0 \\ 0 & a & 1-a \end{bmatrix} \end{array} \begin{array}{c} \begin{matrix} Pr \\ (\text{CR}) \end{matrix} \\ \begin{bmatrix} p_C \\ p_A \\ p_N \end{bmatrix} \end{array} \begin{array}{c} \begin{matrix} Pr \\ (\text{start}) \end{matrix} \\ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{array}. \qquad [1]$$

The rows of the transition matrix represent possible states of responsiveness on any Trial n of an experiment, and the columns represent possible states on Trial $n + 1$. The cells of the matrix give the probability of S moving from one given state to another during the intertrial interval. The entries in the first vector give the probability of a response during the observation interval for each of the three states of responsiveness. The entries in the second vector give the probability of an S starting the experiment in a particular state.

It is assumed that S starts out in an initial, naïve state (N), in which the probability of a response to the CS is P_N. With rabbits, P_N would not be a free parameter, but would take on the a priori value of zero. Following each application of the UCS, there is a fixed probability a that S becomes activated, aroused, or sensitized. This is represented by the process moving from State N to State A. When S is activated, the CS may evoke a response with a fixed probability P_A. After S has become activated, there is a fixed probability c that the response becomes conditioned to the CS. This is represented by the process moving from State A to State C. When conditioning has occurred, the probability is quite high that the CS evokes a blink before the onset of the UCS. This probability is represented by P_C. Thus, the model predicts that there are three discrete levels of performance (P_N, P_A, and P_C) with an all-or-none transition from one level to the next.

MATHEMATICAL ANALYSIS OF THE MARKOV MODEL

In order to make predictions from the Markov model, it will be necessary to specify the probability that S is in a given state on Trial n of an experiment, where $n = 1, 2, 3, \ldots, M$. Let $w_{i,n}$ represent the probability that S is in State i at the beginning of Trial n, where $i = N, A, C$. From the transition matrix in Equation 1 it can be seen immediately that

$$w_{N,n} = (1 - a)^{n-1}. \qquad [2]$$

For an S to be in State A on Trial n he must have remained in State N for k trials ($k = 1, 2, \ldots, n - 1$) before moving to State A, where he remained for $(n-k-1)$ trials. Thus, the probability of being in State A at the beginning of Trial n is

$$w_{A,n} = \sum_{k=1}^{n-1} (1-a)^{k-1} a (1-c)^{n-k-1}. \qquad [3]$$

The solution to the sum in Equation 3 is

$$w_{A,n} = a(c - a)^{-1} \times [(1 - a)^{n-1} - (1 - c)^{n-1}]. \qquad [4]$$

The probability that S is absorbed in the conditioned state, C, by Trial n may be obtained by subtraction as

$$w_{C,n} = 1 - w_{N,n} - w_{A,n}. \qquad [5]$$

The theoretical probability of a CR as a function of trials will be a weighted sum of the probabilities of a CR in each of the three states, where the weights are the probabilities of being in each of the states on Trial n,

$$Pr(\mathrm{CR}_n) = P_N w_{N,n} + P_A w_{A,n} + P_C w_{C,n}. \qquad [6]$$

It can be seen from Equations 2, 4, and 5 that, as the number of trials becomes large, $w_{N,n}$ and $w_{A,n}$ approach limits of zero, and $w_{C,n}$ approaches a limit of unity. Thus,

$$\lim_{n\to\infty} Pr(\mathrm{CR}_n) = P_C. \qquad [7]$$

Probabilities of Response Sequences over Trials n to $n + 3$

The trial-by-trial sequential structure of the data may be assessed by considering sequences of ordered 4-tuples of responses for individual rabbits. Define a random variable X_n which indicates the response made by an S on Trial n:

$$X_n = \begin{cases} 0 & \text{if CR on Trial } n, \\ 1 & \text{otherwise.} \end{cases} \quad [8]$$

Over Trials n to $n + 3$ there are 16 possible response outcome sequences of the form $0_{s,n} = \{X_n, X_{n+1}, X_{n+2}, X_{n+3}\}$ for $s = 1, 2, \ldots, 16$:

$$\begin{array}{ll} 0_{1,n} = \{0000\} & 0_{9,n} = \{1000\} \\ 0_{2,n} = \{0001\} & 0_{10,n} = \{1001\} \\ 0_{3,n} = \{0010\} & 0_{11,n} = \{1010\} \\ 0_{4,n} = \{0011\} & 0_{12,n} = \{1011\} \\ 0_{5,n} = \{0100\} & 0_{13,n} = \{1100\} \\ 0_{6,n} = \{0101\} & 0_{14,n} = \{1101\} \\ 0_{7,n} = \{0110\} & 0_{15,n} = \{1110\} \\ 0_{8,n} = \{0111\} & 0_{16,n} = \{1111\}. \end{array} \quad [9]$$

Given that an S is in a particular state i on Trial n, $i = N, A, C$, the conditional probability of a particular response sequence $0_{s,n}$ may be represented by $Pr(0_s|i)$ which is independent of the trial number n. The $(16\times 3) = 48$ expressions for $Pr(0_s|i)$ may be written down in a straightforward manner from an examination of the transition matrix and response-probability vector in Equation 1. They are listed in the Appendix. The unconditional probability of a particular response sequence beginning on Trial n is a weighted sum of the three conditional probabilities represented by $Pr(0_s|i)$,

$$\begin{aligned} Pr(0_{s,n}) &= w_{N,n}Pr(0_s|N) \\ &\quad + w_{A,n}Pr(0_s|A) \\ &\quad + w_{C,n}Pr(0_s|C). \quad [10] \end{aligned}$$

EVALUATION OF THE MODELS

In evaluating the Markov model, two approaches are available. First it can be assumed that Ss are homogeneous with respect to the values of the parameters of the model. Estimates of P_N may be obtained either from a pretest period in which Ss are given trials of CS alone or from the observed proportion of responses on Trial 1 of the experiment. An estimate of P_C may be taken as the asymptotic proportion of CRs, given that a sufficiently large number of trials has been run. Estimates of the values of the remaining parameters may be obtained by a minimum χ^2 procedure. Let M represent the total number of trials given in an experiment, T the total number of Ss in the experiment, and $F(0_{s,n})$ the observed frequency of Ss who have response sequence $0_{s,n}$ beginning on Trial n. Then,

$$\chi^2 = \sum_{n=1}^{M-3} \sum_{s=1}^{16} [T\cdot Pr(0_{s,n}) - F(0_{s,n})]^2 \times [T\cdot Pr(0_{s,n})]^{-1}. \quad [11]$$

The $Pr(0_{s,n})$, and thus χ^2, are functions of the parameters P_N, a, P_A, c, and P_C. Since estimates of P_N and P_C are directly observable, we need only select values of a, P_A, and c which minimize the χ^2. The minimization can easily be accomplished with a high-speed computer programmed to search for optimum parameter values. This minimum χ^2 estimation technique and its properties have been discussed by Atkinson and Crothers (1964).

It should be noted that the χ^2 defined in Equation 11 is not to be interpreted as the χ^2 as used in hypothesis testing. Independence of observations has not been preserved in Equation 11. Each S in an experiment contributes $M-3$ observations to the χ^2. Since the assumption of independence of observations has been violated, it cannot be guaranteed that the nonindependent χ^2 will have the same sampling distribution as the independent χ^2. Thus, one cannot make a probability statement about a particular value of a nonindependent χ^2 unless its sampling distribution has been determined. The nonindependent χ^2 is, however, an extremely useful statistic for parameter estimation. It may be interpreted as an overall measure of the weighted squared error between a set of theoretical predictions and a set of empirical data. The squared error of each predicted point is weighted by the reciprocal of the theoretically expected number of observations at that point. With respect to theoretical models, the model which predicts the least weighted squared error, that is, yields the minimum nonindependent χ^2, is of course the best fitting model. Appropriate parameter estimates can be obtained for any model by finding those parameter values which yield the minimum weighted squared error for that model.

As an alternative to the assumption that *S*s are homogeneous with respect to the parameter values, it can be assumed that there are individual differences among *S*s with respect to the values of the parameters. The data for individual *S*s would have to be summed over trials and individual parameter values estimated for each *S*. The χ^2 would then become

$$\chi^2 = \sum_{t=1}^{T} \sum_{s=1}^{16} \Big[\sum_{n=1}^{M-3} Pr(0_{s,n,t}) - \sum_{n=1}^{M-3} F(0_{s,n,t})\Big]^2 \times \Big[\sum_{n=1}^{M-3} Pr(0_{s,n,t})\Big]^{-1}, \quad [12]$$

where t is an *S* index, $t = 1, 2, \ldots, T$.

For experiments with a large number of *S*s each one of which is run for many trials, as was the case in the present study, the analysis required to obtain a least χ^2 fit for each *S* is very time-consuming and extremely expensive, even with a very high-speed computer. Thus, from the standpoint of both research cost and theoretical parsimony, it would be desirable if homogeneity could be assumed, allowing predictions to be made with a minimum number of estimated parameters. It follows, therefore, that the model should initially be tested under the assumption of homogeneity; that is, the test in Equation 11 should be conducted first. Only if the weighted squared error is large should the test in Equation 12, with its inherent assumption of heterogeneity of parameter values, be performed.

The χ^2 test requires a reasonably large expected number of observations in each of the response sequence categories. It is often extremely expensive to run enough *S*s to insure a sufficient frequency of all possible response sequences beginning on a particular Trial n. In order to circumvent this problem, it may be necessary to sum over all trials or blocks of trials in the experiment. Under the homogeneity assumption and summing over all trials, the χ^2 becomes

$$\chi^2 = \sum_{s=1}^{16} \Big[T \sum_{n=1}^{M-3} Pr(0_{s,n}) - \sum_{n=1}^{M-3} F(0_{s,n})\Big]^2 \times \Big[T \sum_{n=1}^{M-3} Pr(0_{s,n})\Big]^{-1}. \quad [13]$$

Sequential Analysis for the Response-Independent Models

According to the response-independent models,

$$Pr(X_n = i \mid X_{n-1} = j, X_{n-2} = k, \quad X_{n-3} = m, \ldots) = Pr(X_n = i), \qquad [14]$$

for $i, j, k, m, \ldots = 0, 1$. In the introduction, p_n was used to represent $Pr(X_n = 0)$. Thus, $Pr(X_n = 1)$ is equal to $1 - p_n$ or q_n. It therefore follows that the probability of a particular response sequence $0_{s,n} = \{X_n = i, \quad X_{n+1} = j, \quad X_{n+2} = k, X_{n+3} = m\}$ will be equal to

$$Pr(0_{s,n}) = (q_n)^i (p_n)^{1-i} (q_{n+1})^j (p_{n+1})^{1-j} \cdot (q_{n+2})^k (p_{n+2})^{1-k} (q_{n+3})^m (p_{n+3})^{1-m}, \qquad [15]$$

for $i, j, k, m = 0, 1$. Having theoretical expressions for the $0_{s,n}$, it is then possible to obtain a measure of the weighted squared error of the response-independent model through the χ^2 procedures given in Equations 11, 12, or 13.

ANALYSIS OF THE DATA

The 4-tuple sequences of responses will be analyzed first with respect to the Markov model. We proceed by specifically assuming that the rabbits are homogeneous with respect to parameter values. As mentioned earlier, P_N takes on an a priori value of zero and will not be considered as a free parameter. Every rabbit reached a criterion of 10 consecutive CRs within 150 trials. To estimate P_C, the mean proportions of CRs on trials subsequent to the criterion run of 10 consecutive CRs were computed and appear in Table 1. As can be seen, the proportions are stationary at a value close to .95. Thus, the estimate of P_C will be taken as .95. Estimates of the parameters a, P_A, and c were obtained by having a Control Data Corporation 1604 computer select values which minimized the χ^2 given in Equation 13. The parameter values which minimized the χ^2 were $a = .030$, $P_A = .375$, and $c = .037$. The observed and expected frequencies of the 4-tuple response sequences, summed over all trials of the experiment, are given in Table 2. An examination of Table 2 indicates that the Markov model is following the profile of the data very well and yields a minimum χ^2 of 19.66.

TABLE 1

PROPORTION OF CRs ON TRIALS SUBSEQUENT TO A CRITERION RUN OF 10 CONSECUTIVE CRs

Blocks of 5 trials	Proportion of CRs
1	.93
2	.94
3	.95
4	.94
5	.94
6	.93
7	.95

In spite of the fact that the model is accurately predicting the frequencies of the response sequences, it might be argued that since we have summed over trials, the model may not be faithfully predicting the trial-by-trial changes in the probability of a CR. To assess the trial-by-trial changes in response probability, we require that the Markov model predict the mean learning curve, Equation 6, using the parameter values estimated on the basis of the 4-tuple sequences of responses. The observed and predicted proportions of CRs in blocks of 10 trials are given in Figure 1. It is difficult to imagine how the predictions could be improved. The Markov model is doing an excellent job predicting the trial-

TABLE 2

OBTAINED FREQUENCIES OF 4-TUPLE RESPONSE SEQUENCES AND PREDICTED FREQUENCIES FROM THE MARKOV AND RESPONSE-INDEPENDENT MODELS AND THE SINGLE-OPERATOR LINEAR MODEL WITH INDIVIDUAL DIFFERENCES

Sequence	Data	Markov	Independent	Linear
0000	7361	7348	4779	5663
0001	473	461	973	872
0010	450	468	997	892
0011	176	149	401	367
0100	469	488	1024	914
0101	127	150	415	379
0110	161	162	431	392
0111	231	215	428	395
1000	553	544	1051	941
1001	160	161	433	395
1010	151	173	448	409
1011	217	229	459	423
1100	249	222	467	427
1101	241	253	490	451
1110	326	311	523	481
1111	3502	3513	1528	1446
χ^2		19.66	6422.41	5309.49

by-trial changes in mean response probability.

Consider now the entire class of response-independent learning models which have the property that the occurrence of a response on some Trial n is a function only of the probability of the response on that trial and is independent of the sequence of responses on all preceding trials. An adequate response-independent model would have to predict values of $\{p_1, p_2, p_3, \ldots, p_M\}$ which would accurately describe the obtained mean learning curve given in Figure 1. Under the hypothesis that the subjects are homogeneous, the maximum likelihood estimate of p_n for the most general response independent model is just the proportion of CRs on Trial n. The maximum likelihood estimates of p_n from the rabbit data were entered into Equation 15 to predict the sequences $0_{s,n}$. Using Equation 13, a χ^2 was then computed between the predicted and obtained frequencies of 4-tuples of response sequences, which are given in Table 2. As can be seen, the response-independent predictions are not following the profile of the data. The data sequences are predominantly all CRs or all failures to respond (non-CRs). The response-independent model indiscriminately predicts too many sequences containing both CRs and non-CRs. The χ^2, the measure of the weighted squared error, between obtained and predicted frequencies is extremely large, 6,422.41. It is obvious that the response-independent model is giving an extremely poor approximation to the sequential features of the data.

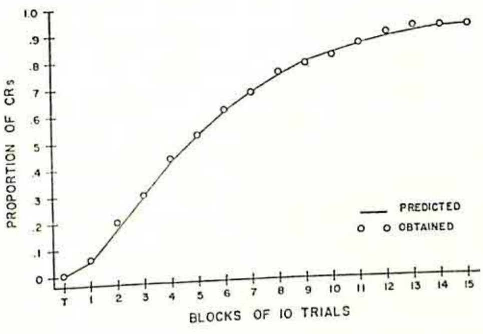

FIG. 1. Obtained group mean learning curve and prediction from the Markov model.

Individual Differences

It is obvious that under the assumption of homogeneity of *S*s, the entire class of response-independent models cannot be considered a realistic approximation to the eye-blink conditioning process. The next logical step is to assume that a response-independent model holds for conditioning in individual *S*s, but that *S*s differ with respect to the values of the parameters of the model. Unfortunately, a reasonable estimate of

p_n for the general case of the response-independent model is not available from the data of a single *S*. The estimate of p_n would be unity or zero depending upon whether the subject made a CR or not, respectively, on Trial *n*. With these estimates, the response-independent model would be trivially true since it would predict the data exactly. It would, however, use up all the degrees of freedom in the data to do so. An alternative approach would be to assume some specific case of the response-independent model which would yield stable parameter estimates for individual subjects. This would leave a finite number of degrees of freedom and would permit meaningful predictions to be made.

The most familiar and most widely applied case of the response-independent model is the single-operator linear model (Bush & Mosteller, 1955; Estes, 1950), which requires that for an individual subject *i*

$$p_{i,n} = \lambda_i - (\lambda_i - p_{i,1})\alpha_i^{n-1}. \quad [16]$$

In Equation 16, $p_{i,n}$ is the theoretical probability of a CR by subject *i* on Trial *n*, λ_i is the asymptotic probability of a CR by subject *i*, and α_i is a rate of growth parameter which can take on only values between zero and unity. For experiments in which the observed proportion of CRs for an individual is asymptotically stable at some value λ_i, the maximum likelihood estimate of the parameter α_i is

$$\hat{\alpha}_i = 1 - (\lambda_i - p_{i,1})/(N_i\lambda_i - S_i), \quad [17]$$

where N_i is the total number of trials subject *i* was run in the experiment, and S_i is the total number of CRs which he made in the N_i trials. To obtain an estimate of α_i, it is first necessary to obtain appropriate estimates of the initial and asymptotic response probabilities. In general, the best way to obtain an estimate of $p_{i,1}$ is to run a pretest in which the CS is given alone for a number of trials. In the experiment under discussion, none of the 101 rabbits blinked to the CS on 10 pretest trials with CS alone, nor did any *S* blink to the CS on the first trial of conditioning. Thus, the maximum likelihood estimate of $p_{i,1}$ for all *S*s is zero. Inspection of the group mean learning curve in Figure 1 indicates that response probability did not change during the last 30 trials of the experiment, Trials 121–150. From the standpoint of the linear model, a case can be made that the data are asymptotic after Trial 120. For each *S* an estimate of λ_i was obtained as the proportion of CRs on Trials 121–150. Taking each *S*'s λ_i estimate and assuming $p_{i,1} = .0$, a maximum likelihood estimate of α_i was obtained for each *S* using Equation 17. Table 3 gives the obtained frequency distribution of the estimates of α_i and λ_i for the single-operator linear model.

TABLE 3

FREQUENCIES OF INDIVIDUAL PARAMETER VALUES FOR THE SINGLE-OPERATOR LINEAR MODEL

Parameter value	α (rate)	λ (asymptote)
under .80	4	6
.80–	1	0
.82–	0	3
.84–	0	10
.86–	1	0
.88–	8	10
.90–	3	0
.92–	8	25
.94–	19	31
.96–	50	0
.98–	6	16

Table 2 presents the 4-tuple sequences of responses predicted by the linear model with individual rate and asymptote parameters. A comparison can be made with the actual data and with the predictions from the Markov and general response-independent models which assume homogeneity of the parameters. The measure of the weighted squared error for the linear model with individual differences is only slightly better than that of the general response-independent model under the homogeneity assumption, $\chi^2 = 5309.49$. Thus, even permitting individual differences in rate and asymptote of learning, the linear model is unable to give a satisfactory account of the third-order sequential dependencies in the data. By comparison, the fit of the Markov model is amazing. A total of 203 parameters were estimated from the entire data matrix of 15,150 responses in order to get the predictions for the linear model with individual differences and 150 parameters were estimated to get the predictions for the general response-independent model, but only 5 parameters estimated to get the predictions for the Markov model.

COMPARISON OF THE MODELS

Given estimates of $p_{i,n}$ it is possible to obtain a likelihood function for the entire *S*s × Trials data matrix of an experiment. For any response-independent model the likelihood of any specific data sequence N_i trials long for a particular subject i is

$$l_i = \prod_{n=1}^{N_i} (1 - p_{i,n})^{X_{i,n}}(p_{i,n})^{1-X_{i,n}}, \quad [18]$$

where $X_{i,n}$ equals 0 if subject i makes a CR on Trial n and equals 1 otherwise. The joint likelihood L over a total of I *S*s is just

$$L = \prod_{i=1}^{I} l_i. \quad [19]$$

To obtain a likelihood function for the Markov model, we need to consider the trial on which S made a transition from the naïve state to the activated state and the trial on which S made a transition from the activated state to the conditioned state. Designate these two trials as j and k where j is less than k. The trials on which the transitions occurred cannot be observed directly from the data, but they have some likelihood of occurring on any trial from 1 to N. For the Markov model it can be shown that the likelihood of the particular response sequence N trials long given by subject i under the hypothesis of homogeneity of the parameter values across *S*s is l_i' where

$$\begin{aligned} l_i' = \sum_{j=1}^{N-2} \sum_{k=j+1}^{N-1} \{ & a(1-a)^{j-1}\Big[\prod_{n=1}^{j} (1-P_1)^{X_{i,n}}(P_1)^{1-X_{i,n}}\Big] \\ & \times c(1-c)^{k-j-1}\Big[\prod_{n=j+1}^{k} (1-P_2)^{X_{i,n}}(P_2)^{1-X_{i,n}}\Big] \\ & \times \Big[\prod_{n=k+1}^{N} (1-P_3)^{X_{i,n}}(P_3)^{1-X_{i,n}}\Big]\} \qquad [20] \\ + \sum_{j=1}^{N-1} \{ & a(1-a)^{j-1}\Big[\prod_{n=1}^{j} (1-P_1)^{X_{i,n}}(P_1)^{1-X_{i,n}}\Big] \\ & \times (1-c)^{N-j-1} \prod_{n=j+1}^{N} (1-P_2)^{X_{i,n}}(P_2)^{1-X_{i,n}}\} \\ + (1-a)^{N-1} & \Big[\prod_{n=1}^{N} (1-P_1)^{X_{i,n}}(P_1)^{1-X_{i,n}}\Big], \end{aligned}$$

As before, the joint likelihood of the entire *S*s × Trials data matrix for the Markov model is just

$$L' = \prod_{i=1}^{I} l_i'. \qquad [21]$$

At this point it should be stressed that a likelihood value is a number between zero and unity which specifies how well a particular model is predicting the data from a particular experiment in *all* its aspects. A likelihood value is sensitive to all the sequential dependencies in the data. By contrast, the χ^2 approach used earlier is sensitive only to third-order sequential dependencies. Thus, the likelihood approach is much more powerful than the χ^2 approach, and inadequate models can be detected more easily with a likelihood approach. When considering a set of alternative models with respect to a specific set of data, the model which yields the largest likelihood value is the model which is predicting the data best and which has the highest probability of being isomorphic with the process which actually generated the data. Thus, all other considerations aside, the model which yields the largest likelihood value is the best fitting model.

Likelihood Ratio Evaluation

The ratio of two likelihoods from two different models also has very interesting properties. Let R represent the ratio of two likelihood values, with the smaller likelihood being placed in the denominator. It is a well-known statistical result that if one of the models is a submodel of the other, $2 \log_e R$ is distributed as χ^2 where the degrees of freedom are equal to the difference in the number of parameters estimated for the numerator (general) model minus the number of parameters estimated for the denominator (restricted) model.

Further, even if the likelihoods are from models of unrelated classes, the ratio R can be expressed as odds in favor of one model over the other. Specifically, the hypothesis that the data were generated by a process isomorphic with the model yielding the higher likelihood is R times more likely of being true than the hypothesis that the data were generated by a process isomorphic with the model yielding the lower likelihood. Thus, the likelihood ratio can be viewed as a measure of the relative efficiency of prediction of one model over another.

Consider, now, the data of the rabbit experiment discussed earlier. For the general response-independent model with 150 estimated parameters, the maximum likelihood L of the entire data matrix of 15,150 trial outcomes is 10^{-3141}. For the single-operator linear model with individual differences the corresponding maximum likelihood is 10^{-2319} with 203 estimated parameters. For the Markov model the likelihood of the entire data matrix given values of the 5 parameters estimated by the minimum χ^2 technique is 10^{-1913}. Comparing the response-independent model with no individual differences with the linear model with individual differences, the likelihood ratio R is 10^{822}. Thus, the linear model with individual differences is predicting the data much closer than the general response-independent model without individual differences. Comparing the Markov model with no individual differences with the single-operator linear model with individual differences, the likelihood ratio R is 10^{406}, in favor of the Markov model. Thus the hypotheses that the Markov model is isomorphic with the conditioning process is 10^{406} times more probable than the hypothesis that the linear model and the conditioning process are isomorphic. This is true in spite of the fact that the Markov model has only 5 estimated parameters while the linear model has 203. These likelihood ratio comparisons give a precise statistical summary of what is obvious from inspection of Table 2. The Markov model is the best fitting of the three models considered in this paper.

THEORETICAL IMPLICATIONS

The analyses indicate that there are definite sequential dependencies in eye-blink responses of individual rabbits. Early in conditioning, non-CRs tend strongly to follow non-CRs. Later, CRs tend strongly to follow CRs. The observed dependence is consistent with the three-state Markov model under the hypothesis that the rabbits are homogeneous with respect to rate of learning and asymptotic level of performance. The observed dependence in the data is, however, too great to be accounted for by a response-independent model such as the single-operator linear model of Bush and Mosteller (1955) and Estes (1950), the single-operator beta model of Luce (1959), and the response-strength model of Spence (1956). It is possible that the latter two models could account for the data if the assumptions of homogeneity of the learning rate and asymptote were relaxed. It remains to be seen, however, whether it is possible with individual parameter estimates to obtain a likelihood value for either of these two models which would be larger than the value yielded by the three-state Markov model. Even if it proves possible, then the question of parsimony must be considered in comparing the models, since the Markov model has only 5 estimated parameters while the models with individual differences will have many more.

In addition to predictive ability and parsimony, the Markov model has the desirable property of generality. It is a member of a family of models with three discrete levels of performance which have been shown to describe accurately learning in many situations with several species of organisms. In humans, for example, Markov models of this type successfully described paired-associates learning (Atkinson & Crothers, 1964), recognition and recall learning (Evans & Dallenbach, 1965; Kintsch & Morris, 1965; Waugh & Smith, 1962), and some aspects of eyelid conditioning (Bower & Theios, 1964). Related Markov models also accurately describe shock avoidance conditioning in rats and dogs (Theios, 1963; Theios & Brelsford, 1966) and shock-motivated choice behavior in rats (Bower & Theios, 1964; Theios, 1965). Thus, on the criteria of precision of prediction, generality, and parsimony, there is reason to continue using and investigating the three-state Markov model as a tool in classical conditioning research.

There are a number of alternative learning models which permit both gradual increases in response probability over trials for individual subjects and large response-dependence effects. These include the two-operator linear model of Bush and Mosteller (1955), the two-operator beta model of Luce (1959), and the two-phase model of Norman (1964). It is possible that these models could account for the response dependencies in the rabbit data. Because of the nature of these models, however, the algebraic task of deriving expressions for the predicted response sequences is quite formidable.

There is no reason why the Markov model considered in this paper should be restricted to eye-blink conditioning studies in rabbits. It is directly applicable to eyelid-conditioning studies involving human *S*s. The only change would be that the parameter P_N would take on a nonzero value and would have to be estimated. The model which Bower and Theios (1964) applied to Gormezano's human eyelid-conditioning data differs from the present model only in that in the present model the initial probability of a CR may be greater than zero and the asymptotic probability of a CR is less than unity. As a result, the model does not require that the proportion of CRs be stationary on the intermediate trials between an *S*'s first CR and his last non-CR, as does the earlier model of Bower and Theios (1964). In fact, the present model predicts that the probability of a CR should increase on these intermediate trials. This follows from the fact that during the intermediate trials *S* may move from the naïve state to the activated state or from the activated state to the conditioned state. Within the naïve, activated, and conditioned states the probability of a response to the CS is assumed to be stationary at values of P_N, P_A, and P_C, respectively. However, if a transition from one state to the next state takes place during the intermediate trials, there will result a consequent increase in the probability of a CR, since P_C is larger than P_A, which is larger than P_N. Thus the model permits increasing CR probabilities on the intermediate trials. The model does predict stationarity of the intermediate responses only when the initial probability of a CR is *absolutely* zero and the asymptotic probability of a CR is *absolutely* unity. Prokasy (1965) has published data from eight eyelid-conditioning studies with human *S*s in which the response probability on the intermediate trials is not stationary. In each study, the initial CR probability was greater than zero and the asymptotic CR probability was less than unity.

Since the present Markov model permits S-shaped mean learning curves with initial levels of responding greater than zero, asymptotes less than 100% responding, and nonstationarity of response probability on trials between the first CR and last non-CR, we know of no published human eyelid-conditioning data which is inconsistent with it. Because of the extremely high variability which is often found in the data of human eyelid-conditioning studies, it will be necessary to take into consideration individual differences among *S*s. These individual differences may take on several forms. For example, *S*s may differ with respect to either learning rate or asymptotic performance level, but be homogeneous with respect to initial level of performance. Alternatively, *S*s may be homogeneous with respect to learning rate and asymptote, but may differ with respect to their initial level of performance. With human *S*s it may not be reasonable to assume that all *S*s start the experiment in the naïve state as was assumed with rabbits. Even if *S*s were homogeneous with respect to the values of the parameters, it is possible that a proportion α of the humans start out in the activated state and a proportion β start out in the conditioned state. Thus, only a proportion $1 - \alpha - \beta$ would start out naïve. If this state of affairs were true, an examination of the data might lead one to the erroneous conclusion that *S*s were heterogeneous with respect to learning *rate*, since some *S*s would give CRs early in training and others not until quite late in training. It would be expected that α and β would be related to such variables as the type of instructions or "set" given to *S*, *S*'s past history, and whether or not a masking task is used.

REFERENCES

ATKINSON, R. C., & CROTHERS, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. *Journal of Mathematical Psychology*, 1964, **1**, 285–315.

BOWER, G. H., & THEIOS, J. A learning model for discrete performance levels. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964. Pp. 1–31.

BRELSFORD, J., JR., & THEIOS, J. Single session conditioning of the nictitating membrane in the rabbit: Effect of intertrial interval. *Psychonomic Science*, 1965, **2**, 81–82.

BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.

DEAUX, E. B., & GORMEZANO, I. Eyeball retraction: Classical conditioning and extinction in the albino rabbit. *Science*, 1963, **141**, 630–631.

ESTES, W. K. Toward a statistical theory of learning. *Psychological Review*, 1950, **57**, 94–107.

EVANS, R. B., & DALLENBACH, K. M. Single-trial learning: A stochastic model for the recall of individual words. *American Journal of Psychology*, 1965, **78**, 545–556.

GORMEZANO, I., SCHNEIDERMAN, N., DEAUX, E., & FUENTES, I. Nictitating membrane: Classical conditioning and extinction in the albino rabbit. *Science*, 1962, **138**, 33–34.

KINTSCH, W., & MORRIS, C. J. Application of a Markov model to free recall and recognition. *Journal of Experimental Psychology*, 1965, **69**, 200–206.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

NORMAN, M. F. A two-phase model and an application to verbal learning. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964. Pp. 173–187.

PROKASY, W. F. Pattern models, sampling theory, $D(A(1-10^{-iN}))$ . . . and some conditioning data. Paper read at Psychonomic Society, Niagara Falls, Ontario, Canada, October 10, 1964.

PROKASY, W. F. An application of a three-state all-or-none model to human eyelid conditioning data. *Psychonomic Science*, 1965, **3**, 349–350.

SCHNEIDERMAN, N., FUENTES, I., & GORMEZANO, I. Acquisition and extinction of the classically conditioned eyelid response in the albino rabbit. *Science*, 1962, **136**, 650–652.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

THEIOS, J. Simple conditioning as two-stage all-or-none learning. *Psychological Review*, 1963, **70**, 403–417.

THEIOS, J. The mathematical structure of reversal learning in a shock-escape T-maze: Overtraining and successive reversals. *Journal of Mathematical Psychology*, 1965, **2**, 26–52.

THEIOS, J., & BRELSFORD, J., JR. Theoretical interpretations of a Markov model for avoidance conditioning. *Journal of Mathematical Psychology*, 1966, **3**, 140–162.

WAUGH, N. C., & SMITH, J. E. K. A stochastic model for free recall. *Psychometrika*, 1962, **27**, 141–154.

APPENDIX

CONDITIONAL PROBABILITIES of SEQUENCES

As defined in the text, $Pr(0_s|i)$ represents the conditional probability of a specific response sequence $0_{s,n}$ given that the conditioning process is in State i of the Markov chain on Trial n, where $i = N, A, C$. Let $Q_i = (1 - P_i)$ and

$$B = (1 - a)P_N$$
$$F = (1 - a)Q_N$$
$$Z = (1 - c)P_A$$
$$Y = (1 - c)Q_A.$$

Then:

$$Pr(0_1|N) = P_N[B^3 + B^2aP_A + BaP_A(Z + cP_C) + aP_A(Z^2 + ZcP_C + cP_C^2)]$$
$$Pr(0_1|A) = P_A(Z^3 + Z^2cP_C + ZcP_C^2 + cP_C^3)$$
$$Pr(0_1|C) = P_C^4$$
$$Pr(0_2|N) = P_N[B^2F + B^2aQ_A + BaP_A(Y + cQ_C) + aP_A(ZY + ZcQ_C + cP_CQ_C)]$$
$$Pr(0_2|A) = P_A(Z^2Y + Z^2cQ_C + ZcP_CQ_C + cP_C^2Q_C)$$
$$Pr(0_2|C) = P_C^3Q_C$$
$$Pr(0_3|N) = P_N[B^2F + BFaP_A + BaQ_A(Z + cP_C) + aP_A(ZY + YcP_C + cP_CQ_C)]$$
$$Pr(0_3|A) = P_A(Z^2Y + ZYcP_C + ZcP_CQ_C + cQ_CP_C^2)$$
$$Pr(0_3|C) = P_C^3Q_C$$
$$Pr(0_4|N) = P_N[BF^2 + BFaQ_A + BaQ_A(Y + cQ_C) + aP_A(Y^2 + YaQ_C + cQ_C^2)]$$
$$Pr(0_4|A) = P_A(ZY^2 + ZYcQ_C + ZcQ_C^2 + cQ_C^2P_C)$$
$$Pr(0_4|C) = P_C^2Q_C^2$$
$$Pr(0_5|N) = P_N[B^2F + FBaP_A + FaP_A(Z + cP_C) + aQ_A(Z^2 + ZcP_C + cP_C^2)]$$
$$Pr(0_5|A) = P_A(YZ^2 + YZcP_C + ZcP_C^2 + cP_C^2Q_C)$$
$$Pr(0_5|C) = P_C^3Q_C$$
$$Pr(0_6|N) = P_N[BF^2 + BFaQ_A + FaP_A(Y + cQ_C) + aQ_A(ZY + ZcQ_C + cP_CQ_C)]$$
$$Pr(0_6|A) = P_A(Y^2Z + YZcQ_C + YcP_CQ_C + cQ_C^2P_C)$$
$$Pr(0_6|C) = P_C^2Q_C^2$$
$$Pr(0_7|N) = P_N[F^2B + F^2aP_A + FaQ_A(Z + cP_C) + aQ_A(ZY + YcP_C + cP_CQ_C)]$$
$$Pr(0_7|A) = P_A(Y^2Z + ZYcP_C + YcP_CQ_C + cP_CQ_C^2)$$
$$Pr(0_7|C) = P_C^2Q_C^2$$
$$Pr(0_8|N) = P_N[F^3 + F^2aQ_A + FaQ_A(Y + cQ_C) + aQ_A(Y^2 + YcQ_C + cQ_C^2)]$$
$$Pr(0_8|A) = P_A(Y^3 + Y^2cQ_C + YcQ_C^2 + cQ_C^3)$$
$$Pr(0_8|C) = P_CQ_C^3$$

For $s = 9, 10, 11, \ldots, 16$, a general equation for $Pr(0_s|i)$ is as follows:

$$Pr(0_s|i) = R_iPr(0_{s-8}|i),$$

where $R_i = Q_i/P_i$ and $i = N, A, C$.

TABLE 4

CODED SEQUENTIAL DATA FOR INDIVIDUAL RABBITS (*S*)

S	Length of successive runs of non-CRs (odd positions) and CRs (even positions)
1	19 1 6 4 2 1 1 6 1 36 1 10 1 2 1 49 1 8
2	37 1 3 2 1 1 1 4 1 1 1 3 1 93
3	15 1 2 1 7 1 15 2 3 3 1 2 1 3 1 4 1 16 1 5 1 32 1 17 1 13
4	27 1 1 2 3 1 1 19 1 31 1 16 1 11 1 10 1 22
5	19 1 11 1 1 1 4 5 1 4 2 1 1 24 1 3 1 2 1 5 1 12 1 47
6	32 1 4 1 2 1 12 1 2 2 5 1 1 7 1 2 8 2 2 7 1 4 1 1 1 8 1 3 2 5 1 1 1 16 1 9
7	11 1 16 1 65 1 3 1 3 1 8 1 2 1 2 1 1 1 7 1 1 1 1 4 1 2 1 8 1 10
8	2 1 11 1 1 1 3 1 4 3 1 8 1 2 1 6 1 22 1 23 1 3 1 7 1 26 1 16
9	19 1 1 1 4 5 1 44 1 26 1 5 1 11 1 8 1 1 1 3 1 13
10	3 1 3 1 8 1 2 4 2 1 2 2 1 5 2 6 1 2 1 4 1 21 1 70 1 1 1 2
11	32 1 1 1 9 1 16 2 3 1 1 3 2 4 1 3 1 1 1 1 7 1 1 1 1 3 3 17 1 2 1 23 1 3
12	42 2 5 5 1 2 1 2 1 2 1 4 2 2 1 3 2 5 2 5 1 4 1 4 1 28 1 20
13	2 1 54 1 1 2 1 1 1 3 3 1 1 5 2 9 1 1 1 4 2 2 1 6 1 5 1 18 1 15 1 2
14	6 2 3 1 4 1 4 1 1 1 1 2 17 1 14 2 6 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 4 1 3 1 1 4 1 14 1 17 1 16
15	12 1 1 1 2 13 1 20 1 97 1
16	78 1 2 1 2 4 3 4 2 13 1 15 1 23
17	22 1 1 1 6 1 1 1 1 1 1 2 1 3 1 2 2 2 1 2 2 2 2 5 1 38 1 2 1 7 1 16 1 13 1 4
18	33 1 18 1 1 2 3 1 3 1 11 4 6 4 1 7 4 21 1 4 1 7 1 9 1 2 1 1
19	44 1 10 1 13 1 2 3 1 1 1 1 3 1 4 1 3 1 14 1 1 1 1 1 1 1 3 3 4 1 1 1 10 1 2 2 4 5
20	24 1 5 1 6 1 7 1 5 1 2 2 2 3 2 4 3 3 2 3 1 6 2 8 2 13 1 9 1 29
21	51 1 3 2 3 2 3 2 2 1 3 44 1 17 1 24
22	19 1 6 2 4 1 1 2 13 1 8 1 7 1 2 2 3 1 3 2 21 5 2 13 1 28
23	5 1 4 1 2 2 3 2 3 1 2 3 2 2 3 4 4 1 2 1 2 13 1 1 1 25 1 37 1 2 1 3 1 13
24	8 1 11 2 7 2 2 3 10 3 8 53 1 30 1 8
25	31 1 4 2 11 2 7 1 4 2 2 4 3 5 2 5 2 3 1 58
26	39 1 5 2 2 3 1 3 2 2 2 12 1 56 1 17 1
27	48 1 8 2 1 3 3 3 2 3 2 2 8 2 2 4 2 40 1 4 1 5 1 2
28	9 1 3 3 1 24 1 5 1 27 1 12 1 28 1 32
29	3 2 3 4 3 5 2 6 2 10 1 94 1 14
30	56 1 8 1 2 1 3 4 2 5 2 5 2 5 3 50
31	38 1 6 1 2 3 2 3 1 3 11 2 2 2 3 1 1 12 1 45 1 9
32	11 2 3 33 1 24 1 52 1 16 1 5
33	15 1 1 1 13 1 3 2 9 1 3 1 1 1 1 6 1 10 1 3 1 50 1 1 1 11
34	2 23 1 28 1 12 1 13 1 23 2 2 1 29 1 6 1 3
35	49 2 7 2 5 1 1 46 1 23 1 6 1 5
36	53 3 2 2 1 4 1 1 1 1 3 42 1 17 1 7
37	95 1 1 1 4 1 2 1 1 5 1 2 1 2 1 3 1 13 2 12
38	27 38 1 4 1 22 1 2 1 33 1 19
39	27 1 4 1 2 1 2 50 1 5 1 4 1 9 1 20 1 16 1 2
40	40 1 1 2 3 1 2 93 1 1 2 3
41	25 3 1 1 11 24 1 13 1 7 1 3 1 18 1 3 1 35
42	7 2 1 85 1 9 1 10 1 10 2 21
43	16 2 1 35 3 9 1 1 1 80 1
44	9 2 1 5 1 2 1 1 2 20 1 13 1 1 1 59 1 11 1 17
45	14 16 1 6 1 10 1 16 1 20 1 18 1 9 1 15 1 18
46	19 1 2 48 1 31 1 6 1 5 1 16 1 3 1 13
47	5 16 1 2 1 45 1 47 1 11 1 1 1 17
48	23 1 1 1 2 10 1 14 1 61 1 34
49	19 1 3 1 3 1 1 2 1 27 1 31 1 19 1 1 1 36
50	29 1 2 11 1 21 1 7 2 3 1 13 1 4 1 9 1 15 1 1 1 11 1 5 1 6
51	1 1 6 40 1 52 1 8 1 39
52	35 3 2 1 1 1 1 3 6 1 4 3 2 1 1 2 1 31 1 2 2 44 1 1
53	21 1 1 1 1 1 2 1 4 2 1 10 1 27 1 22 28 1 23
54	16 3 1 41 1 17 1 44 1 2 1 9 1 3 1 8

Table 4—Continued

S	Length of successive runs of non-CRs (odd positions) and CRs (even positions)
55	75 2 5 1 2 1 7 1 4 1 1 2 1 1 1 45
56	35 1 10 1 5 1 7 1 2 1 3 1 1 8 1 35 1 17 1 10 1 2 1 3 1
57	24 1 7 1 2 28 1 27 1 53 1 4
58	15 1 1 4 1 19 1 24 1 10 1 14 1 20 1 36
59	38 1 11 4 3 1 2 2 2 3 1 5 2 1 1 1 1 38 1 29 1 1 1
60	14 28 1 3 1 17 1 29 1 18 1 1 1 1 1 11 2 11 1 7
61	20 1 4 5 1 7 1 16 1 94
62	14 1 6 38 1 19 1 7 1 19 1 5 1 26 1 7 1 1
63	5 1 1 22 1 7 1 12 2 45 1 16 1 2 1 30 1 1
64	12 2 4 2 4 4 2 25 1 1 1 25 1 8 1 7 1 49
65	11 11 1 4 1 31 1 5 1 36 1 14 1 9 1 17 1 4
66	38 2 2 21 1 3 1 2 1 9 1 25 1 12 1 10 1 1 1 4 1 72
67	17 2 6 1 1 1 7 1 2 2 1 1 1 1 2 2 1 1 1 3 1 2 1 2 2 3 1 1 1 7 7 1 1 1 1 10 1 24 1 28
68	1 2 1 1 1 3 2 2 2 3 1 2 3 3 1 24 1 3 1 1 1 1 1 2 1 3 1 20 1 17 1 11 1 4 1 17 1 8
69	4 19 1 7 1 15 1 12 1 1 1 1 1 13 1 9 1 7 1 25 1 7 1 19
70	7 2 1 1 1 2 1 35 1 99
71	7 1 32 1 5 1 1 1 9 2 8 3 3 1 3 3 8 2 5 2 4 1 1 1 1 2 1 26 1 14
72	28 1 1 1 1 1 1 1 2 1 2 2 2 1 1 1 3 7 1 12 1 20 1 58
73	18 1 2 3 1 1 2 1 2 1 1 24 1 11 1 18 1 30 1 14 1 7 1 2 1 3 1
74	45 1 5 2 1 1 5 2 5 4 1 1 5 21 1 13 2 19 1 6 1 1 1 5 1
75	58 1 1 1 1 1 2 14 1 4 1 2 2 30 1 30
76	74 2 3 1 2 3 3 1 1 1 1 4 1 1 1 3 2 1 2 2 1 4 1 2 1 2 2 2 1 1 1 1 1 3 1 1 2 1 1 1 5 10
77	49 1 5 1 1 1 6 2 2 1 1 1 1 3 2 20 1 11 1 8 1 3 1 2 1 24
78	56 1 2 1 4 2 2 1 2 1 4 2 2 1 1 1 4 63
79	81 1 3 1 3 2 7 1 4 1 1 1 2 1 7 1 2 1 1 2 2 25
80	26 1 3 1 1 20 1 23 1 37 1 10 1 5 2 10 1 6
81	9 3 2 2 3 2 5 1 1 1 2 2 3 1 2 18 1 21 1 22 1 29 1 17
82	20 1 1 1 4 1 30 1 6 2 2 3 1 4 1 6 8 1 1 1 3 2 1 1 1 1 11 2 2 3 3 7 1 1 1 4 1 1 1 2 1 4 1 10
83	5 1 2 1 3 1 1 1 1 2 8 1 1 2 3 1 1 1 1 1 1 1 1 2 3 1 2 3 1 2 1 4 3 1 5 11 1 6 1 41 1 10 2 8
84	42 75 1 19 1 12
85	70 3 4 2 1 1 1 1 1 1 2 5 1 1 1 2 1 3 1 2 1 1 1 1 1 22 1 2 1 6 1 2 1 5
86	31 1 3 1 7 1 4 2 3 1 1 3 1 2 1 46 1 5 1 9 1 25
87	26 2 8 1 2 1 6 3 2 3 1 1 1 14 1 15 1 15 1 29 1 3 1 12
88	3 1 1 43 2 22 1 12 1 28 1 33 1 1
89	23 2 2 1 1 1 4 1 3 84 1 9 1 17
90	10 1 1 1 1 2 1 1 2 3 1 2 4 16 1 24 1 15 2 16 1 10 1 14 1 11 1 6
91	43 1 3 2 3 1 4 2 1 1 3 2 3 3 3 23 2 11 1 21 2 6 1 2 1 5
92	59 1 3 1 6 2 5 1 2 3 4 4 2 1 2 1 1 1 1 1 3 35 1 7 1 2
93	11 2 1 3 2 2 2 1 1 1 2 80 1 17 1 23
94	48 1 1 1 3 1 2 2 5 1 2 2 4 1 4 1 2 2 2 1 3 2 4 2 2 1 3 1 3 1 2 2 1 4 2 3 3 18 1 1 1 4
95	37 1 5 1 2 1 1 1 2 15 1 2 1 15 1 22 1 26 1 4 1 9
96	24 2 1 1 6 41 1 24 1 37 1 3 2 6
97	29 1 1 2 1 1 2 1 3 1 26 1 1 1 1 1 9 1 7 1 2 1 2 2 2 39 2 9
98	101 1 6 1 3 2 1 1 3 1 3 16 1 10
99	113 1 1 11 1 23
100	64 1 7 1 2 6 32 1 1 1 1 1 5 1 7 1 1 1 7 1 5 1 2 10
101	114 1 28 1 6 10

(Received August 4, 1965)

Psychological Review
1966, Vol. 73, No. 5, 409–426

ATTACHMENT BEHAVIOR OF MAMMALS[1]

ROBERT B. CAIRNS

Indiana University

Available data support the empirical generalization that animals tend to remain in the presence of objects to which they have been continuously exposed. From such observations, and certain assumptions of associative learning theory, it is proposed that mammalian attachment behavior is determined by: (a) the length of association with an object in a given context, and (b) the relative cue weight of the object. The model emphasizes the associative nature of the attachment process, the contextual and response relativity of attachment behavior, and the role of the immediate stimulus setting in the control of an animal's response subsystems.

This paper is concerned with the mechanisms by which the social attachments of mammals are acquired, maintained, and extinguished. Following a brief overview of the problem, a preliminary formulation on the nature of the attachment process will be outlined, and the relevant evidence reviewed. To adumbrate the general proposal, no essential distinction is drawn between social and nonsocial attachments, and both classes of behavior are seen as inevitable outcomes of an associative conditioning process.

AN OVERVIEW

Animals tend to approach and remain with others of their kind. This response disposition is distinctive in its persistence and intensity, comparable in such respects to behavior patterns that are immediately relevant to physiological need states of the animal. But despite the prominence of this response tendency, and its obvious relevance to theories of social behavior, there have been until recently few systematic attempts to explore its determinants in mammals (Beach & Jaynes, 1954; Scott, 1962).

Consequently, explanations of social attachment behavior in psychology traditionally have been linked more directly to prevailing general theoretical systems than to the results of specific empirical investigations. Early accounts which emphasized the operation of innate neurophysiological mechanisms, such as the "gregarious instinct" (McDougall, 1908) and "herd instinct" (Trotter, 1916), thus gave way to secondary drive concepts, for example, "drive for gregariousness" (Dollard & Miller, 1950) and dependency motivation (Whiting & Child, 1953). Subsequent studies of the ontogeny of social attachment behavior have, however, yielded meager support for key assumptions of the learned drive model (e.g., Harlow & Zimmermann, 1959; Schaffer & Emerson, 1964; Sears, Maccoby, & Levin, 1957).

Various alternative hypotheses recently have been offered to account for the role that learning plays in the formation of mammalian attachments. In general, social bonds have been presumed to form when an animal, (S) is maintained in close proximity to an object, (O_a) and some crucial condition, or set of conditions, is fulfilled. The latter has referred to characteristics of O_a (e.g., Harlow & Zimmermann, 1959), to particular states of *S* (e.g., Collias, 1962; Scott, 1962), and whether O_a has served as a general source of reinforcement events for *S* (e.g., Gewirtz, 1961).

---

[1] Preparation of this manuscript was facilitated by a fellowship period at The Jackson Laboratory (PHS Special Fellowship 1-F3-MH-30, 205-10). I am indebted to several of my colleagues at Indiana University for critically reading earlier versions of this paper.

But the history of research in this area, brief though it be, indicates that social attachments can occur in the absence of assorted conditions that have been assumed to be primary. Attachments apparently develop in circumstances which do *not* include: lactation by O_a (Harlow, 1958; Igel & Calvin, 1960); immaturity of *S* (Altmann, 1963; Jensen, 1965); physical contact between *S* and O_a (Cairns & Johnson, 1965); or nonpunitiveness of O_a with respect to *S* (Fisher, 1955; Rosenblum & Harlow, 1963; Seay, Alexander, & Harlow, 1964). According to one reviewer (Scott, 1963), available evidence supports the conclusion that various conditions facilitate, but are not necessary for, the formation of social attachments. These results raise the compelling possibility that no special event, environmental or organismic, is an essential accompaniment of proximity. Or, stated positively, the evidence suggests that an animal will form an attachment with respect to any object, animate or inanimate, with which it has been maintained in a proximate relationship (Scott, 1963).

Finally, some primary features of the reference phenomenon should be noted. In the first place, attachments occur discriminatively, for example, animals characteristically become more strongly attached to animate than to inanimate objects. Moreover, the closest social ties typically are formed with respect to one or more conspecific animals. Second, a number of conditions apparently facilitate the attachment formation process. As noted above, recent studies have implicated both organismic and object/situational variables. Third, social bonds are not only formed and maintained, they ordinarily diminish and are extinguished. Though few experimental studies are available with respect to the last point, naturalistic and observational evidence indicates that filial attachments invariably diminish in strength or shift to alternative objects. Any model which attempts to account for the attachment formation process in mammals should, as a minimum, specify the mechanisms which control the selectivity, rapidity of development, and extinguishability of these responses.

TOWARD A THEORY OF MAMMALIAN ATTACHMENT BEHAVIOR

Previous results support the empirical generalization that an animal forms an attachment, of some strength, with respect to those objects with which it has been continuously maintained. Can a case be made for the stronger proposition that attachment strength is determined primarily by the length and closeness of association? Although this hypothesis has parsimony in its favor, it encounters some significant problems. One critical question concerns the selectivity of attachments: Why is it that stronger social ties are formed with respect to animate than inanimate objects? Nor is it clear how a simple associational interpretation of the attachment process would account for the facilitative effects of such variables as lactation or tactile stimulation. Then there is the matter of primary theoretical import: Why should contiguity qua contiguity determine the strength and the direction of social attachment behavior?

Such issues require that explicit attention be given to the nature of the learning process involved in social attachment formation. Following a brief overview of certain assumptions of contiguity learning theory as it has been developed by E. R. Guthrie (1952, 1959) and W. K. Estes (1950, 1958, 1959), some implications of these assumptions for social attachment behavior will be examined.

1. It has been assumed that a stimulus pattern which accompanies a response, and/or immediately precedes it, becomes a cue for that response (Guthrie, 1952; Voeks, 1950). Responses, then, become associated to configurations of stimuli considered as units. Should the response be extended over time, it would be conditioned to the several stimulus patterns sampled during that period.

2. A second assumption concerns the weight, or importance, that any object or environmental event will be assigned in a particular stimulus pattern. In general, it has been assumed that the weight of a subset of stimulus elements is a joint function of the sampling probability and the number of elements within the set (Estes, 1958). Thus, the larger the set, or the greater the sampling probability of its elements, the more important should be its role as a determinant of behavior.

While theorems relating size of a subset to observable properties have yet to be formally developed, it has been assumed that such stimulus aspects as intensity, vividness, and "attention-getting" characteristics covary with the number of elements in the subset (Atkinson & Estes, 1963, p. 202). Whether an event is salient at any given moment would presumably also be a function of its physical involvement in *S*'s ongoing response (cf., the "discriminanda" and "manipulanda" properties of stimuli, Tolman, 1932). Holding sampling probability constant, then, highly conspicuous or response-involved events would be assigned a greater weight than inconspicuous or noninvolved events in the stimulus patterns available to *S*.

3. A third assumption has to do with the development and disruption of response sequences. The integration of ongoing behavior has been accounted for by assuming the formation of response chains, such that the performance of one phase of an activity provides the stimulus components for the next succeeding phase. Stimulus patterns supporting ongoing behavior thus would be made up of internal (e.g., "movement produced," Guthrie, 1952) and external events.

It seems reasonable to expect, furthermore, that a continuously present, heavily weighted cue can become a significant component in a large proportion of the stimulus patterns available to *S*. While such events are usually internal, under some conditions (e.g., continuous and exclusive exposure), prominent external cues would become conditioned to a considerable proportion of the subsystems in the animal's response repertoire. Their presence in stimulus patterns would thus provide for the maintenance of orderly be-

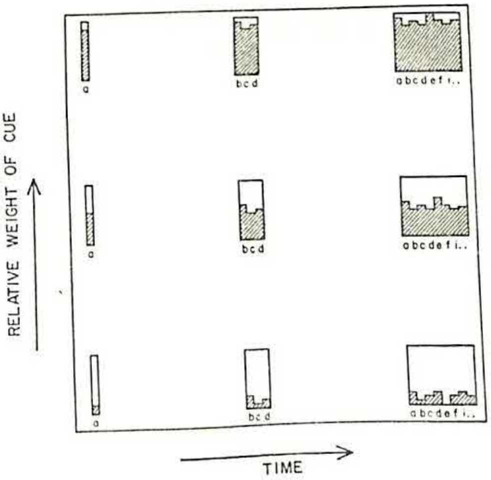

FIG. 1. Schematic representation of the proposed relationship between length of association and cue weight in the conditioning of response subsystems.

havior; their abrupt removal from the context would be the occasion for the disruption of orderly behavioral sequences (cf., "behavior supports," Tolman, 1932, p. 85).

To summarize the foregoing, the probability that an external event will acquire effective control over a given response system of the organism will be determined by the joint operation of two variables: the weight of the stimulus subset, and the extent to which it has been present during the performance of several variations of the response system. This relationship is shown graphically in Figure 1. The ordinate refers to the weight of the stimulus subset relative to other elements in the stimulus pattern; and the abscissa refers to the proportion of stimulus patterns in which the subset is represented. In terms of observable variables, subset weight would vary with stimulus salience (i.e., intensity, movement, etc.) and response-involvement; while proportion of stimulus patterns (corresponding to response subsystems a, b, . . . i, .) would be a function of the length of time that the event was present. The higher the ratio of shaded/total area within any given subrectangle, the greater should be the likelihood that a particular response system would be disrupted by the event's removal.

The integration of the shaded areas across subrectangles should provide an index of the generalized control that any object has over *S*'s behavior. For instance, if a heavily weighted event is represented in all of the stimulus patterns which have been conditioned to *S*'s basic response systems, it would attain generalized control over *S*'s responses. This state of affairs is represented schematically in the upper right-hand corner of Figure 1. The event's removal should be associated with at least a temporary disruption of all of the response systems for which it was relevant.

Following Guthrie (1952), it can further be argued that the reintroduction of a widely conditioned cue (as indicated in upper right of Figure 1) constitutes a "reinforcing" state of affairs. The reasoning can be summarized as follows: (*a*) If the removal of a heavily weighted, generalized cue, C, disrupts all established behavior patterns, then (*b*) a series of disjointed responses, R_1, R_2, R_3, R_4, . . R_s, R_t, will occur in the presence of the remaining stimulus patterns, $S_*^{\bar{\mathrm{c}}}$, and (*c*) each new response in the series, for example, R_4, will alienate its predecessors R_1, R_2, R_3 from their association with $S_*^{\bar{\mathrm{c}}}$. But let some response in the series, R_s, remove the animal from $S_*^{\bar{\mathrm{c}}}$. Then, on subsequent occurrences of $S_*^{\bar{\mathrm{c}}}$, R_s would be evoked. Or, more generally, (*d*) any response, R_s, which reintroduces the widely conditioned cue, or equivalently, removes $S_*^{\bar{\mathrm{c}}}$, will remain conditioned to $S_*^{\bar{\mathrm{c}}}$.

In terms of the present discussion, $S_*^{\bar{\mathrm{c}}}$ thus refers to the class of stimulus patterns produced if an extensively conditioned event were removed abruptly from the total setting. The event's removal would, in effect, make unavailable the population of stimulus patterns to which the animal's basic response systems had been conditioned. The remaining stimulus patterns—those in which the event was not represented—thus would be unable to support the previously established response systems. Any response that removes $S_*^{\bar{\mathrm{c}}}$ would become conditioned to $S_*^{\bar{\mathrm{c}}}$ (cf., Guthrie, 1952, p. 297). In effect, stimulus patterns created by the withdrawal of an extensively conditioned event would be expected, initially, to evoke responses which lead to their own removal.

The above analysis would apply only

in the special instance where an event had become an essential stimulus component in a high proportion of *S*'s response systems. But what would be the case if an event of low relative cue weight were removed? According to the general formulation, the less heavily weighted the cue, the less disruptive would be its removal. The absence of this event presumably would not materially affect the individual stimulus configurations of which it was a part, and the corresponding response systems would reflect little, if any, generalized disruption (Figure 1, lower left and right). Since the remaining stimulus patterns could support previously organized response sequences, these stimulus patterns would not be expected to become selectively conditioned to response sequences which reintroduce the event.

Still a third major possibility can be considered: namely, where an event is heavily weighted in the stimulus patterns associated with certain response systems, but has a low relative weight, or is not represented at all, in other stimulus patterns (Figure 1, upper left). Its removal may or may not lead to behavioral disruption, dependent upon the stimulus patterns being sampled at the time the event was removed.

Implications for Attachment Behavior

The major thesis of this paper can be stated succinctly. It is proposed that an object, or class of objects, can acquire an essential cue function for the maintenance of the response patterns of an animal. Objects which appear in recurrent combination with other environmental or internal events can become significant components in the stimulus patterns which support *S*'s behavior, including such basic maintenance response systems as drinking, reposing, eating, etc. When O_a is present, the associative chains remain intact and stable responding is observed; when O_a is absent, the associative chains are disrupted, and nonorganized responding occurs. At a gross descriptive level, *S* would appear to have formed an attachment with respect to O_a.

Social and nonsocial attachment-objects. In accord with stimulus theory, no qualitative distinction is drawn between the various sources of stimulation. But events clearly differ in terms of such factors as attention-getting properties and potentiality for becoming involved directly in the responses of *S*. Specifically, animals are, among other things, typically more mobile, conspicuous, and capable of developing patterns of reciprocal stimulation than nonliving objects. And it would follow that cues provided by animals are more likely to be observed, and conditioned, than those of inanimate objects. On the other hand, it should be possible to enhance the likelihood of attachment formation (i.e., generalized conditioning) with respect to an inanimate object by manipulating its stimulus characteristics. For instance, if an object is highly conspicuous relative to all other objects in *S*'s maintenance environment, it would be expected to acquire significant cue properties for *S*'s behavior. Or the maintenance environment itself may acquire such properties.

Contextual and response relativity. While the major concern in the present analysis will be with the case where an event is conditioned to a wide range of responses over various contexts, it should be noted that attachment formation may be response- or context-specific. Indeed, it is entirely possible for one object to be conditioned to one set of responses, and another object to be conditioned to an alternative set. Whether multiple relationships would be formed would depend, in part, upon the number of different contexts (objects) in which S has been maintained, and the extent to which distinctive response subsystems have been conditioned to the different contexts.

Attachment maintenance and extinction. The model is obviously not a static one: The question of attachment maintenance is as relevant as the question of attachment acquisition. Should S's basic response systems become conditioned to $S_{*}^{\bar{o}_{a}}$, O_a's removal would produce little if any disruption. It would follow that the longer the period of separation from O_a, the greater would be the likelihood that O_a's presence was not required for the maintenance of general stable responding. On the other hand, those habits that S had acquired specifically with respect to O_a presumably would remain relatively intact if S and O_a were abruptly separated and kept apart. (Note how this state of affairs would be schematized in the upper left of Figure 1.)

Though a detailed comparison of this proposal to other concepts of "social drive" is beyond the scope of this paper, one comment seems in order. In contrast with most biological drive systems, where the number of drive cues is positively correlated with hours of deprivation, it can be argued that the effects of "social" deprivation are at full strength immediately upon separation from O_a. Since reconditioning is presumed to begin immediately upon separation, the extent to which $S_{*}^{\bar{o}_{a}}$ disrupts organized responding should diminish rather than increase as a function of time.

"Punishment" and "rewards." Neither noxious nor positive stimulation is assumed to operate directly upon attachment "strength" (i.e., extent to which O_a is conditioned to S's response subsystems), although both may operate upon it indirectly. In naturalistic settings, noxious stimulation by O_a with respect to S is associated with at least temporary physical separation. To the extent that the separation is extended, attachment strength would be expected to diminish. On the other hand, under some environmental circumstances (e.g., infancy, enforced cohabitation), separation is not possible. To the extent that noxious stimulation from O_a did not inhibit S's orientation toward or the performance of responses with respect to O_a, "punishment" could enhance the salience of O_a and the essential conditioning process.

Whereas noxious stimulation might operate to either increase or decrease attachment strength, such events as the delivery of consumables by O_a would be expected to enhance it. Any condition which makes O_a more prominent, or requires S to interact with O_a, should increase the proportion of stimulus patterns in which O_a is represented. If certain consummatory responses, for instance, invariably occurred in the presence of O_a, these responses would become conditioned to stimulus configurations in which O_a was represented. In general, however, the model indicates that the number of interactions between S and O_a, and not the quality of the interactions or their outcomes, is the immediate determinant of attachment strength.

Level of analysis. Not the least important of the model's implications are those having to do with methodological matters. To evaluate adequately the deductions that follow from it, a detailed analysis of molar response patterns is required, similar to that advocated by Bindra (1961) and Logan (1960) for the study of response systems of the individual, and by Sears (1951) and Schneirla (1946) for the study of social behavior. The problem for research becomes one of specifying the factors in the immediate environment that are controlling an observed response sequence. Because typically two or more animals are involved, the analysis must be concerned with chains of reciprocal or dyadic stimulation, where the behaviors of one animal become the cues for the responses of the other. Even the label "attachment behavior" is seen to be misleading when the disparate responses that are subsumed by this construct are analyzed in terms of their component elements.

Finally, it should be clear that the model is concerned with the analysis of behavioral phenomena (e.g., whether *S* approaches and remains in the presence of another animal), and not with emotional states that have been sometimes presumed to accompany or underlie the behavior (e.g., affection, love, fear, anxiety).

To summarize the association hypothesis rather drastically, it is the proposal that the development and maintenance of intraspecific attachments can be explained as a by-product of a continuing conditioning process. The formation of complex response chains is facilitated by the participation of relatively durable extraorganism events, and it is to these cues that response patterns become "attached." The cues provided by animate objects, by virtue of their salience and involvement in *S*'s behavior, are particularly likely to become conditioned.

Deductions regarding the conditions for the formation and maintenance of social attachments include the following:

1. An attachment of some strength will be formed with respect to those objects with which *S* has been continuously maintained in a proximate relationship.
2. The stimulus characteristics of the object will constitute a significant determinant of the conditioning process, such that the greater its salience and its relevance for *S*'s activities, the more likely that *S*'s response subsystems will become conditioned to the object's presence.
3. Removal of an object which has become conditioned to a high proportion of the response systems of *S* should be associated with a significant disruption of *S*'s ongoing behavior; replacement of the object should be associated with a cessation of disruption.
4. If a generalized conditioned environmental event, O_a, is removed, and some activity of *S* reintroduces it, then this activity is likely to be repeated upon subsequent occasions that O_a is removed. In general, the more complete the behavioral disruption of *S* upon removal of O_a, the greater will be the likelihood that *S* will perform a response that reintroduces O_a.
5. The strength of *S*'s attachment with respect to O_a should be inversely related to the length of separation from O_a.

SOME RECENT EXPERIMENTAL FINDINGS

Within the past 10 years a sufficient number of studies of mammalian attachment behavior have been completed to permit a preliminary evaluation of the association model. Unfortunately for comparison purposes, there has been considerable variation among investigators with respect to the procedures followed and the response measures employed. Diversity of method is not without compensations, however. Since the work has not been focalized exclusively on a single species, it is possible to assess the adequacy of the model in terms of the attachment behavior of several orders of the class *Mammalia*.[2]

---

[2] Only those studies which are seen as being immediately relevant to the association

Effects of Extended Exposure

A primary implication of the association hypothesis is that a prominent environmental event will acquire significant cue properties for an animal which has been maintained in a proximate relationship to it. Data relevant to this expectation are provided by three types of studies: (*a*) interspecific cohabitation; (*b*) isolation; and (*c*) intraspecific cohabitation.

Interspecific cohabitation. The procedure typically followed in these studies has been to rear a young animal with an alien species or distinctive inanimate object, and to assess at some later time the animal's preferences relative to that species or object. Occasionally the interspecific attachment formation of adult *S*s has been studied (e.g., Hersher, Richmond, & Moore, 1963b). Work has been conducted with rodents (e.g., Denenberg, Hudgens, & Zarrow, 1964; Nagy, 1965), carnivores (e.g., Igel & Calvin, 1960; Kuo, 1930, 1960), ungulates (e.g., Cairns, 1964; Hersher, et al., 1963b), and subhuman primates (e.g., Harlow, 1961; Harlow & Zimmermann, 1959; Mason & Green, 1962). Various measures have been employed to assess attachment strength, ranging from indices of time spent in contact with the cohabitant to assessments of choice behavior in a U maze where the subject is permitted to select either his cohabitant or a conspecific animal. The results of these investigations indicate that the experimental animal tends to reinstate the "social" object to which he has been continuously exposed. Cohabitant preference has been demonstrated even when *S* must choose between an alien species and his own (e.g., Cairns & Johnson, 1965; Denenberg, et al., 1964).[3]

Isolation studies. Within the framework of the present theoretical analysis, the isolation experiment is viewed in terms of the cues that are present and available to be conditioned in the "isolation" context. The emphasis is not upon what the animal has been "deprived" of, but what he has been exposed to. This orientation can be contrasted with the tendency to view isolation in terms of its "deprivation" aspects, that is, that which has been withheld from the animal (cf., Lorenz, 1965). Furthermore, it seems paradoxical to expect that an intact animal can be entirely isolated from all of the cues that are distinctive to his species. Certain conspecific events would likely be produced by the animal himself (Schneirla, 1959), for example, elimination odors, self-visual cues, self-grooming stimulation. Despite this reservation, it would be consistent with the association model to expect that animals reared in isolation would be less likely to develop response patterns that require the presence of a conspecific object than would animals which have been maintained with their own species.

In several of the isolation studies that have been conducted, a response index has been used that yields some information about the strength of conspecific preferences. There appears a consistent tendency for animals maintained in isolation, compared with nonisolated *S*s, to show a diminished tendency to approach other members of

---

model have been cited. The review thus is selective rather than comprehensive. Due to space restrictions, discussions of particular methodological issues have been omitted.

[3] An interesting variation of the interspecific cohabitation technique has shown that female mice reared with "perfumed" parents select perfumed male sex-partners more frequently than do females reared with nonperfumed parents. For male *S*'s, however, sexual receptivity of the female apparently overrides the effect of perfume (Mainardi, Marsan, & Pasquali, 1965).

their own species. As in the cohabitation studies, evidence for this generalization is available for various mammalian species, including rats (Angermeier, 1960; Ashida, 1964), cats (Kuo, 1960; Schneirla & Rosenblatt, 1961), sheep (Scott, 1945), and monkeys (Harlow, 1962; Mason, 1960; Sackett, Porter, & Holmes, 1965). These results are generally consistent with the interpretation that the conspecific object may constitute a disruptive cue, or at the least, not be required for the performance of the isolated animal's basic response repertoire.

Intraspecific cohabitation. Information on the determinants of conspecific attachment formation is available for subjects ranging from human and infrahuman primates (e.g., Ainsworth, 1963; DeVore, 1963) to maternal laboratory rats (e.g., Rosenblatt & Lehrman, 1963). With a few exceptions (e.g., Mason, 1961; Nowlis, 1941), the research has focused upon the mother-infant relationship. Intraspecific studies cannot be easily interpreted as being peculiarly relevant to the association assumption, since length of exposure is confounded with several other potentially effective variables, including effects of caretaking, lactation, and possible innate determinants of preference. This work has yielded two general findings, however, that seem immediately relevant to the present hypothesis.

One finding concerns the typical insulation of mother-young units from competing associations (Beach & Jaynes, 1954). For slow-maturing mammals (analogous to altricial birds) insulation is assured by the incompetence of the neonate. Moreover, where the young are precocious, the environment is typically arranged to keep the mother and her offspring isolated. In the case of wild ungulates, for instance, the maternal animal physically separates her infant and herself from the herd (e.g., Altmann, 1963; Darling, 1937). Among domesticated ungulates, further, a functional isolation of maternal-infant pairs is typically maintained by the rejection of the young animal by potential surrogates. Should the neonatal-maternal association be disrupted shortly after parturition, the mutual attachment process is retarded or inhibited (Collias, 1956; Hersher, Richmond, & Moore, 1958).

Second, it appears that prolonged and extensive interaction between a maternal animal and her offspring promotes a dyadic conditioning process, where the behaviors of the one acquire a significant cue function for the other. For instance, Rosenblatt & Lehrman (1963) have observed how the cues of the neonatal rat pup provide the stimuli for the maternal activities of licking and retrieving. Maternal licking, in turn, apparently elicits and directs the neonate's elimination and sucking responses. This behavior of the infant then stimulates further consummatory behavior and stimulation by the mother. Similar descriptions of the reciprocal dependency of maternal-infant behaviors have been given for cats (Schneirla, Rosenblatt, & Tobach, 1963), sheep (Collias, 1956), and, at a more global level, human infants (Schaffer & Emerson, 1964).

While maternal-infant interactions are uniquely suited for the development of a wide range of interresponse contingencies, the primary phenomenon may be also observed in nonfilial relationships. Nowlis (1941) has observed that an important determinant of whether a juvenile chimpanzee prefers one partner over another is *S*'s ongoing behavior: one partner was preferred for mutual grooming; another for play-fighting, etc. (also Mason, 1960). These observations are consistent with the proposal that social preferences reflect the extent to which stimuli provided by an object have become integrated into, and provide cues for, particular response sequences of *S*.

Stimulus Properties of Cohabitant

One of the principal assumptions of the present model is that the likelihood that a given response or set of responses will become conditioned to an environmental event is a direct function of the stimulus weight of that event during the performance of the response. As discussed above, the stimulus weight of the event is presumed to be determined by its perceptual salience and its properties as a manipulanda.

Given this assumption, along with the fact that particular variables facilitate attachment formation (e.g., lactation, tactile stimulation), it might be argued that these stimulus characteristics are effective because they enhance the salience or interactive capabilities of the object. While the claim may be plausible, it can easily become tautological. Unless there is some means to specify a priori the effect that a given variable will have upon the stimulus weight of an event, there seems little gain in asserting that all variables which facilitate attachment behavior are effective because they make the event relatively prominent or become involved directly in *S*'s response patterns. An independent specification of stimulus weight is obviously called for.

It has been generally assumed that the attention-getting properties of an event depend, at least initially, upon such stimulus characteristics as complexity, variability, and intensity. There is some support for this assumption. One series of studies has demonstrated, for example, that rhesus monkeys spend more time viewing varied, moving, or chromatic scenes than nonvaried, static, achromatic ones (Butler, 1961; Butler & Woolpy, 1963). Several investigations have reported, in addition, that complexity is an important determinant of attention in young humans (e.g., Cantor, Cantor, & Ditrichs, 1963; Fantz, 1961; Spears, 1964).

Few attempts have been made, however, to relate stimulus conspicuousness or prominence directly to the strength of subsequent attachments. In a preliminary investigation of this relationship, it was found that lambs tended to form attachments with respect to the salient inanimate objects (i.e., operating television sets) with which they had been individually confined (Cairns, in press, Exp. 1). By the end of the 9th week of cohabitation, the lambs (*a*) became disrupted when the sets were removed, and (*b*) learned to turn in the appropriate direction to reach the inanimate object in a series of U-maze trials. Lambs exposed to dogs for the same period, but debarred from physical interaction with them, developed relatively stronger attachments with respect to their cohabitants than did lambs paired with the inanimate objects.[4]

The above results also support the assumption that environmental events do not have to serve as manipulanda in order to become significant components of the stimulus patterns that support *S*'s behavior. Distance receptor stimuli may play a more important role in mammalian behavioral development than has been generally conceded (e.g., Fuller & Clark, 1966a; Meier, 1965;

---

[4] Supplementary evidence relevant to the stimulus weight-attachment assumption has been reported in recent studies of avian imprinting. Bateson (1964b) found that conspicuousness of a stimulus is a significant determinant of whether chicks will approach it (see also Bateson, 1964a, and Collins, 1965).

Sackett, 1965). Walters and Parke (1965) review related evidence.

Behavioral Disruption

According to the present model, the removal of a previously conditioned subset of stimulus elements, or the introduction of an unconditioned subset, would be expected to interrupt ongoing behavior. The amount and duration of the disruption would reflect, in part, the stimulus weight of the subset and the number of response systems that have been conditioned to it. Stimulus subsets may refer, of course, to a variety of internal or external events, including animals, objects, or places.

Several recent studies have been concerned with the effects of removal from the home environment on behavioral organization. Scott and his collaborators (e.g., Elliot & Scott, 1961; Fredericson, 1950; Ross, Scott, Cherner, & Denenberg, 1960; Scott & Bronson, 1964) have shown that the placement of a dog 3–12 weeks of age alone in an unfamiliar context is associated with heightened activity, yelping, and general behavioral disturbance. The disruptive effects of removal are lessened if (*a*) the animal is placed with another dog in the unfamiliar context, or (*b*) is left alone in the home compartment rather than alone in the unfamiliar setting. Similarly, removal of an interspecific or intraspecific cohabitant from the home compartment has been shown to be associated with behavioral disruption in young lambs (Cairns & Johnson, 1965; Hersher, et al., 1963b) and monkeys (e.g., Jensen & Tolman, 1962). Behavioral disturbance in the human infant upon separation from the mother has been observed in naturalistic (Schaffer & Emerson, 1964) and laboratory contexts (unpublished study conducted by Don E. Fleener in the writer's laboratory).

The marked tendency for the maternal animal to become disrupted when separated from her own or foster offspring also has been reported. Study of attachment development in the pig-tailed monkey indicates that the amount of maternal disruption observed increases as a function of the mother's experience with the neonate (Jensen, 1965). Parallel findings have been reported for interspecific attachment development in maternal ewes and does (Hersher, Richmond, & Moore, 1963a). Moreover, the extent of disruption shown by the subhuman primate mother during separation is negatively correlated with the degree of sensory access that she has to the infant (Jensen & Tolman, 1962).

There have been few attempts to relate the extensity of behavioral disruption of *S* to the tendency of *S* to reinstate the missing object, though such a relationship has been frequently assumed (e.g., Jensen, 1965). In one study of this contingency, behavioral disturbance upon cohabitant removal was found to be a good predictor of U-maze choice behavior (Cairns, in press). Lambs that were most disturbed upon separation from dog cohabitants tended to select these objects rather than conspecific animals. Data are not yet available on the corresponding relationship in other mammalian species.

Effects of Continued Separation

Short-term separation. The present model proposes that separation from an attachment-object initiates a relearning process. If this be the case, then length of the separation interval should be inversely related to the amount of behavioral disruption observed. Where several different intervals of "social deprivation" (i.e., hours since separation from cohabitant) have been sampled, subjects (lambs) show consistently less disruption as a function of time (Cairns, in press, Exp. 2). The greatest decrement in disruption (as measured by activity level and vocalization) occurred in the first 4 hours of separation, with a gradual decline thereafter toward preseparation levels of activity. A similar trend, but over a much shorter period (i.e., 10 minutes) was found for young dogs which had been repeatedly placed in an unfamiliar context (Elliot & Scott, 1961). If the period of isolation is extended over several hours, the disturbance functon obtained for 30 pups drawn from the same age-breed population studied by Elliot and Scott (1961) is quite similar to that observed in ungulates (unpublished data of the writer).

One other line of evidence requires comment. The studies of Stanley and his collaborators (Bacon & Stanley, 1963; Stanley & Elliot, 1962; Stanley, Morris, & Trattner, 1965) have been viewed as being pertinent to the problem of social "satiation." Dogs which have been exposed to the testing context, or to some significant portion of it (maze-human, or either maze or human separately), immediately prior to a test trial, tend to run slower than if preexposure had not occurred. It seems reasonable to suggest, along the lines indicated by Stanley (1965), that the testing context itself changes in cue function, probably as a joint function of the number of repeated trials and the intertrial interval. Further study of the "satiation" phenomenon might include exploration of the effects of different intervals of prolonged habitation in the testing context (i.e., a procedure analogous to extended cohabitation).

Extended separation. Few studies have been directly concerned with the effects of extended separation upon a previously formed attachment. The data that are available indicate that prolonged separation from an original cohabitant, and placement with a different animal, leads to a reversal of the original preference in lambs (Cairns & Johnson, 1965; Cairns, in press, Exp. 3). Although the reversal occurred more rapidly when the "new" cohabitant was of the same species as *S*, the essential reversal occurred when the "new" O_a was a dog (Cairns, in press).

Studies of the postemergence attachment behavior of animals reared in isolation are also relevant. Although there have been several studies of the effects of isolation, few investigators have been concerned with the influence of subsequent experience on the attachment behavior of isolated animals. In addition, where reports have appeared, the tendency has been to highlight the persisting effects of the treatment rather than the ameliorative influence of the postisolation experience.

Many of the effects of isolation appear, in fact, to be at least partially reversible (e.g., Fuller, 1966; Fuller & Clark, 1966b; Meier, 1965; Seay, Alexander, & Harlow, 1964). In a review of primate studies, Mason & Riopelle (1964) suggest that the severity of the consequences of social isolation is roughly proportional to the length and amount of restriction. More work is called for on the mechanisms by which these effects are mediated. It seems not unreasonable to suggest, in this regard, that the context in which *S* is placed may be partly responsible for the persistence of "asocial" responses. For example, a lamb which has been removed from the maternal ewe is rejected by all ewes, at least temporarily, when reintroduced to the flock. The reported failure of isolated ungulates to develop "normal" conspecific relations may be as much an outcome of the flock's response to *S* as the converse (Hersher, Richmond, & Moore, 1958; Scott, 1945).

Punishment and separation. Social attachments are not necessarily inhibited or extinguished by punishment. Attachments form even when *S* is bitten, mauled, or otherwise abused by O_a (Cairns & Johnson, 1965; Rosenblum & Harlow, 1963; Seay, Alexander, & Harlow, 1964). These observations were made, however, under special conditions of environmental restriction: Neither *S* nor O_a were permitted to escape, and alternative objects were not present.

A different outcome is obtained under less restrictive circumstances. In naturalistic or semi-naturalistic settings, continued punishment is typically associated with physical separation and extinction of the social attachment (e.g., Altmann, 1960). Whether punishment is an antecedent or a consequent of attachment diminution unfortunately cannot be determined from field studies.

Other reports suggest that maturation events initiate chains of reciprocal responding which culminate in punishment and attachment extinction. If, for instance, suckling responses persist after a given stage of bone and musculature development, they presumably are noxious and elicit maternal punishment and avoidance (Schneirla & Rosenblatt, 1961; Rheingold, 1963). Similarly, young dogs cohabited individually with an adult rabbit tend to sleep beside or atop him, and to spend most of the time that they are awake oriented toward him, for example, tugging his tail, mouthing his ears, scratching his neck. If the pups are young (28–32 days old), the rabbits remain passive and sometimes become involved in reciprocal grooming. By the time the dogs are 60–70 days old, however, the picture of peaceful coexistence is inverted. Responses of the pups that were once harmless now inflict serious physical damage and elicit vigorous avoidance and defensive biting by the rabbit (unpublished study conducted by the writer and J. Werboff).

Moreover, it would be inappropriate to focus only upon changes in the offspring. Since the maturation of the young is usually synchronized with maternal physiological changes, it seems reasonable to assume tentatively that the occurrence of punishment is determined by modifications in the biophysical status of both.

Finally, punishment in itself would not be expected to diminish attachment strength. According to the present analysis, noxious stimulation would influence mutual attachments only if the animals move or orient away from each other. Insofar as available data are pertinent, they seem consistent with the proposal.

Organismic Variables

Age of subject. Most of the work on attachment behavior has been conducted with young animals as subjects. But whatever mechanisms are responsible for the development of social attachment behavior in the young apparently are not entirely eliminated as the animal matures. Maternal females, at least, demonstrate a similar capacity in adulthood. Some of the factors that may facilitate mutual mother-young reciprocal conditioning have already been discussed: the physical or functional isolation of mother-young units from competing associations, the continuous involvement of one animal across a large number of the response systems of the other. In addition, certain age-related structural and physiological changes that might diminish attachment strength and preclude substitutive attachments have been noted. Other potentially effective variables, such as maturational-experiential changes in sensory encoding and transmission, have yet to be adequately explored in terms of their effects upon attachment formation and extinction. While such developmental changes doubtless affect the essential conditioning process, it would be premature to assign differential weights to the presumably effective variables. That task can best await the accumulation of a larger body of relevant data than is now available.

On the other hand, the empirical analysis of the ontogenesis of attachment phenomena would likely be facilitated by appropriate assumptions about the nature of the process itself. It is with respect to this matter that the working assumption of the present model can be underscored. All social attachments, regardless of the object (mother, peer, mate, offspring, etc.) or the age of *S* (infancy, juvenile, adult) are presumed to reflect a single process; namely, the associative conditioning of *S*'s response subsystems to the cues provided by O's presence.

Genetic variation. In view of the differences that obtain between the various mammalian orders in terms of structure, breeding and rearing patterns, and presumed capabilities for the establishment of conditioned responses, it would be surprising if corresponding interspecific differences in attachment behavior could not be identified. Whether these differences can be shown to be a direct function of phylogenetic rank is, of course, another matter. Pertinent to this issue are the findings of recent studies of "social reinforcement" phenomena in rodents, that is, where a mature *S* is permitted to view a conspecific animal if *S* performs a particular response. Relative to other outcomes, social reinforcement events provide only modest control over the behavior of rodents (e.g., Allison & Rocha e Silva, 1965; Angermeier, 1962; Antonitis & Baron, 1964; Antonitis & Kish, 1955). These results stand in sharp contrast to findings obtained in similar test arrangements with "higher" mammalian species (e.g., ungulates, primates). Without additional comparative study, it is unclear whether the apparent discrepancy reflects the operation of genetic, age, or outcome condition variables.

Cross-specific differences would not necessarily be in conflict with the present model. Its utility in this regard may lie in the specification of the mechanisms whereby between-breed or between-species differences are likely to be mediated. For example, the dominant receptor systems in infancy or adulthood would be expected to be the primary determinants of the stimulus weight that a given environmental variable would be assigned. Once reliable interspecific attachment behavior differences have been established, investigation of the variables that mediate the differences would be in order.

CONCLUDING REMARKS

The task of this paper was to develop an integrated and coherent account of social attachment behavior in mammals. Within the limits that have been set for it, the association hypothesis is consistent with a considerable range of data, and provides a reasonably systematic account of the processes of attachment acquisition, maintenance, and extinction. But the model is obviously incomplete. Some of its limitations have been duly noted in the preceding pages. While the formulation possibly can be extended to deal with the problems raised by genetic and individual differences, such an extension must be deferred until more complete data become available.

Finally, a comment on the general orientation of the model. By emphasizing the role of the immediate stimulus setting in the control of behavior, it underscores the need for a detailed analysis of molar response patterns, on the one hand, and the factors in the stimulus setting that support these responses, on the other. The model thus forces attention onto the mechanisms by which behavioral stability is achieved, and the processes by which the effects of prior experience are mediated. If a microanalysis of molar responses can clarify the amorphous construct of "general activity" (Bindra, 1961), it might be able to perform a similar service for the social behavior rubrics of "dependency" and "attachment."

REFERENCES

AINSWORTH, M. The development of infant-mother interaction among the Ganda. In B. M. Foss (Ed.), *Determinants of infant behavior: II.* New York: Wiley, 1963. Pp. 67–104.

ALLISON, J., & ROCHA E SILVA, M. Time spent with food and nonfood incentives as a function of food deprivation. *Psychonomic Science,* 1965, **2**, 63–64.

ALTMANN, M. The role of juvenile elk and moose in the social dynamics of their species. *Zoologica,* 1960, **45**, 35–39.

ALTMANN, M. Naturalistic studies of maternal care in moose and elk. In H. L. Rheingold (Ed.), *Maternal behavior in mammals.* New York: Wiley, 1963. Pp. 233–253.

ANGERMEIER, W. F. Some basic aspects of social reinforcements in albino rats. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 364–367.

ANGERMEIER, W. F. The effect of a novel and novel-noxious stimulus upon social operant behavior in the rat. *Journal of Genetic Psychology,* 1962, **100**, 151–154.

ANTONITIS, J. J., & BARON, A. Social reinforcement of bar pressing in mice. *Journal of Genetic Psychology,* 1964, **105**, 223–236.

ANTONITIS, J. J., & KISH, G. B. Reactions of C57 black male mice to active and inactive social stimuli. *Journal of Genetic Psychology,* 1955, **86**, 115–130.

ASHIDA, S. Modification by early experience of the tendency toward gregariousness in rats. *Psychonomic Science,* 1964, **1**, 343–344.

ATKINSON, R. C., & ESTES, W. K. Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. G. Galanter (Eds.), *Handbook of mathematical psychology: II.* New York: Wiley, 1963. Pp. 121–268.

BACON, W. E., & STANLEY, W. C. Effect of deprivation level in puppies on performance maintained by a passive person reinforcer. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 783–785.

BATESON, P. P. G. Effect of similarity between rearing and testing conditions on chicks' following and avoidance responses. *Journal of Comparative and Physiological Psychology,* 1964, **57**, 100–103. (a)

BATESON, P. P. G. Relation between conspicuousness of stimuli and their effectiveness in the imprinting situation. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 407–411. (b)

BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychological Bulletin,* 1954, **51**, 239–263.

BINDRA, D. Components of general activity and the analysis of behavior. *Psychological Review,* 1961, **68**, 205–215.

BUTLER, R. A. The responsiveness of rhesus monkeys to motion pictures. *Journal of Genetic Psychology,* 1961, **98**, 239–245.

BUTLER, R. A., & WOOLPY, J. H. Visual attention in the rhesus monkey. *Journal of Comparative and Physiological Psychology,* 1963, **56**, 324–328.

CAIRNS, R. B. Social behavior changes by cross-species rearing. *American Psychologist,* 1964, **19**, 484. (Abstract)

CAIRNS, R. B. Development, maintenance, and extinction of social attachment behavior in sheep. *Journal of Comparative and Physiological Psychology,* 1966, **62**, in press.

CAIRNS, R. B., & JOHNSON, D. L. The development of interspecies social preferences. *Psychonomic Science,* 1965, **2**, 337–338.

CANTOR, G. N., CANTOR, J. H., & DITRICHS, R. Observing behavior in preschool children as a function of stimulus complexity. *Child Development,* 1963, **34**, 683–689.

COLLIAS, N. E. The analysis of socialization in sheep and goats. *Ecology,* 1956, **37**, 228–239.

COLLIAS, N. E. Social development in birds and mammals. In E. L. Bliss (Ed.), *Roots of behavior.* New York: Harper, 1962. Pp. 264–273.

COLLINS, T. B., JR. Strength of the following response in the chick in relation to degree of "parent" contact. *Journal of Com-*

parative and Physiological Psychology, 1965, **60**, 192–195.

DARLING, F. F. *A herd of red deer.* London: Oxford University, 1937.

DENENBERG, V. H., HUDGENS, G. A., & ZARROW, M. X. Mice reared with rats: Modification of behavior by early experience with another species. *Science,* 1964, **143**, 380–381.

DEVORE, I. Mother-infant relations in free-ranging baboons. In H. L. Rheingold (Ed.), *Maternal behavior in mammals.* New York: Wiley, 1963. Pp. 305–335.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

ELLIOT, O., & SCOTT, J. P. The development of emotional distress reactions to separation, in puppies. *Journal of Genetic Psychology,* 1961, **99**, 3–22.

ESTES, W. K. Toward a statistical theory of learning. *Psychological Review,* 1950, **57**, 94–107.

ESTES, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: University of Nebraska Press, 1958. Pp. 35–69.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 380–491.

FANTZ, R. L. The origin of form perception. *Scientific American,* 1961, **204**, 66–72.

FISHER, A. E. The effects of differential early treatment on the social and exploratory behavior of puppies. Unpublished doctoral dissertation, Pennsylvania State University, 1955.

FREDERICSON, E. Distributed versus massed experience in a traumatic situation. *Journal of Abnormal and Social Psychology,* 1950, **45**, 259–266.

FULLER, J. L. Transitory effects of experiential deprivation upon reversal learning in dogs. *Psychonomic Science,* 1966, **4**, 273–274.

FULLER, J. L., & CLARK, L. D. Effects of rearing with specific stimuli upon post-isolation behavior in dogs. *Journal of Comparative and Physiological Psychology,* 1966, **61**, 258–263. (a)

FULLER, J. L., & CLARK, L. D. Genetic and treatment factors modifying the postisolation syndrome in dogs. *Journal of Comparative and Physiological Psychology,* 1966, **61**, 251–257. (b)

GEWIRTZ, J. L. A learning analysis of the effects of normal stimulation, privation, and deprivation on the acquisition of social motivation and attachment. In B. M. Foss (Ed.), *Determinants of infant behavior.* New York: Wiley, 1961. Pp. 213–290.

GUTHRIE, E. R. *The psychology of learning.* (2nd ed.) New York: Harper, 1952.

GUTHRIE, E. R. Association by contiguity. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 158–195.

HARLOW, H. F. The nature of love. *American Psychologist,* 1958, **13**, 673–685.

HARLOW, H. F. The development of affectional patterns in infant monkeys. In B. M. Foss (Ed.), *Determinants of infant behavior.* New York: Wiley, 1961. Pp. 75–88.

HARLOW, H. F. The heterosexual affectional system in monkeys. *American Psychologist,* 1962, **17**, 1–9.

HARLOW, H. F., & ZIMMERMANN, R. R. Affectional responses in the infant monkey. *Science,* 1959, **130**, 421–432.

HERSHER, L., RICHMOND, J. B., & MOORE, A. U. Effect of post partum separation of mother and kid on maternal care in the domestic goat. *Science,* 1958, **128**, 1342–1343.

HERSHER, L., RICHMOND, J. B., & MOORE, A. U. Maternal behavior in sheep and goats. In H. L. Rheingold (Ed.), *Maternal behavior in mammals.* New York: Wiley, 1963. Pp. 203–232. (a)

HERSHER, L., RICHMOND, J. B., & MOORE, A. U. Modifiability of the critical period for the development of maternal behavior in sheep and goats. *Behaviour,* 1963, **20**, 311–320. (b)

IGEL, G. J., & CALVIN, A. D. The development of affectional responses in infant dogs. *Journal of Comparative and Phyiological Psychology,* 1960, **53**, 302–305.

JENSEN, G. D. Mother-infant relationship in the monkey *Macaca nemestrina:* Development of specificity of maternal response to own infant. *Journal of Comparative and Physiological Psychology,* 1965, **59**, 305–308.

JENSEN, G. D., & TOLMAN, C. W. Activity level of the mother monkey, *Macaca nemestrina,* as affected by various conditions of sensory access to the infant following separation. *Animal Behaviour,* 1962, **10**, 228–230.

KUO, Z. Y. The genesis of the cat's responses to the rat. *Journal of Comparative Psychology,* 1930, **11**, 1–35.

KUO, Z. Y. Studies on the basic factors in animal fighting: VII. Interspecies coexistence in mammals. *Journal of Genetic Psychology,* 1960, **97**, 211–225.

LOGAN, F. A. *Incentive: How the conditions of reinforcement affect the performance of rats.* New Haven: Yale University Press, 1960.

LORENZ, K. *Evolution and modification of behavior.* Chicago: University of Chicago Press, 1965.

MAINARDI, D., MARSAN, M., & PASQUALI, A. Causation of sexual preferences of the house mouse. The behaviour of mice reared by parents whose odour was artificially altered. *Atti della Società Italiana di Scienze Naturali e del Museo Civico di Storia Naturale di Milano,* 1965, **104**, 325–338.

MASON, W. A. The effects of social restriction on the behavior of rhesus monkeys: I. Free social behavior. *Journal of Comparative and Physiological Psychology,* 1960, **53**, 582–589.

MASON, W. A. The effects of social restriction on the behavior of rhesus monkeys: II. Tests of gregariousness. *Journal of Comparative and Physiological Psychology,* 1961, **54**, 287–290.

MASON, W. A., & GREEN, P. C. The effects of social restriction on the behavior of rhesus monkeys. IV. Responses to a novel environment and to an alien species. *Journal of Comparative and Physiological Psychology,* 1962, **55**, 363–368.

MASON, W. A., & RIOPELLE, A. J. Comparative psychology. *Annual Review of Psychology,* 1964, **15**, 143–180.

McDOUGALL, W. *An introduction to social psychology.* London: Methuen, 1908.

MEIER, G. W. Other data on the effects of social isolation during rearing upon adult reproductive behaviour in the rhesus monkey (*Macaca-mulatta*). *Animal Behaviour,* 1965, **13**, 228–231.

NAGY, Z. M. Effect of early environment upon later social preference in two species of mice. *Journal of Comparative and Physiological Psychology,* 1965, **60**, 98–101.

NOWLIS, V. Companionship preference and dominance in the social interaction of young chimpanzees. *Comparative Psychological Monographs,* 1941, **17**, (1, Whole No. 85).

RHEINGOLD, H. L. Maternal behavior in the dog. In H. L. Rheingold (Ed.), *Maternal behavior in mammals.* New York: Wiley, 1963. Pp. 169–202.

ROSENBLATT, J. S., & LEHRMAN, D. S. Maternal behavior of the laboratory rat. In H. L. Rheingold (Ed.), *Maternal behavior in mammals.* New York: Wiley, 1963. Pp. 8–57.

ROSENBLUM, L. A., & HARLOW, H. F. Approach-avoidance conflict in the mother-surrogate situation. *Psychological Reports,* 1963, **12**, 83–85.

ROSS, S., SCOTT, J. P., CHERNER, M., & DENENBERG, V. H. Effects of restraint and isolation on yelping in puppies. *Animal Behaviour,* 1960, **8**, 1–5.

SACKETT, G. P. Effects of rearing conditions upon the behavior of rhesus monkeys (*Macaca mulatta*). *Child Development,* 1965, **36**, 855–868.

SACKETT, G. P., PORTER, M., & HOLMES, H. Choice behavior in rhesus monkeys: An effect of stimulation during the first month of life. *Science,* 1965, **147**, 304–306.

SCHAFFER, H. R., & EMERSON, P. E. The development of social attachments in infancy. *Monographs of the Society for Research in Child Development,* 1964, **29**, (3, Whole No. 94).

SCHNEIRLA, T. C. Problems in the biopsychology of social organization. *Journal of Abnormal and Social Psychology,* 1946, **41**, 385–402.

SCHNEIRLA, T. C. An evolutionary and developmental theory of biphasic processes underlying approach and withdrawal. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1959.* Lincoln: University of Nebraska Press, 1959. Pp. 1–42.

SCHNEIRLA, T. C., & ROSENBLATT, J. S. Behavioral organization and genesis of the social bond in insects and mammals. *American Journal of Orthopsychiatry,* 1961, **31**, 223–253.

SCHNEIRLA, T. C., ROSENBLATT, J. S., & TOBACH, E. Maternal behavior in the cat. In H. L. Rheingold (Ed.), *Maternal behavior in mammals.* New York: Wiley, 1963. Pp. 122–168.

SCOTT, J. P. Social behavior, organization and leadership in a small flock of domestic sheep. *Comparative Psychological Monographs,* 1945, **18**, (4, Whole No. 96).

SCOTT, J. P. Critical periods in behavioral development. *Science,* 1962, **138**, 949–958.

SCOTT, J. P. The process of primary socialization in canine and human infants. *Monographs of the Society for Research in Child Development,* 1963, **28**, (1, Whole No. 85).

SCOTT, J. P., & BRONSON, F. H. Experimental exploration of the et-epimeletic or care-soliciting behavioral system. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior.*

Stanford: Stanford University Press, 1964. Pp. 174–193.

SEARS, R. R. A theoretical framework for personality and social behavior. *American Psychologist*, 1951, **6**, 476–483.

SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing*. New York: Harper, 1957.

SEAY, B., ALEXANDER, B. K., & HARLOW, H. F. Maternal behavior of socially deprived rhesus monkeys. *Journal of Abnormal and Social Psychology*, 1964, **69**, 345–354.

SPEARS, W. C. Assessment of visual preference and discrimination in the four-month-old infant. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 381–386.

STANLEY, W. C. The passive person as a reinforcer in isolated beagle puppies. *Psychonomic Science*, 1965, **2**, 21–22.

STANLEY, W. C., & ELLIOT, O. Differential human handling as reinforcing events and as treatments influencing later social behavior in basenji puppies. *Psychological Reports*, 1962, **10**, 775–788.

STANLEY, W. C., MORRIS, D. D., & TRATTNER, A. Conditioning with a passive person reinforcer and extinction in Shetland sheep dog puppies. *Psychonomic Science*, 1965, **2**, 19–20.

TOLMAN, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.

TROTTER, W. *Instincts of the herd in peace and war*. London: Fisher Unwin, 1916.

VOEKS, V. W. Formalization and clarification of a theory of learning. *Journal of Psychology*, 1950, **30**, 341–362.

WALTERS, R. H., & PARKE, R. D. The role of the distance receptors in the development of social responsiveness. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior*. Vol. 2. New York: Academic, 1965. Pp. 59–96.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale University Press, 1953.

(Received June 7, 1965)

Psychological Review
1966, Vol. 73, No. 5, 427-436

TESTS FOR SPONTANEOUS ALTERNATION[1]

ERIC W. HOLMAN[2]

University of Pennsylvania

Predictions about sequential properties of 2-choice learning data are derived from a simple assumption about the effects of rewarded and unrewarded choices. These predictions are used as tests for spontaneous alternation in T-maze experiments. Given massed trials, rats show spontaneous alternation during position and reversal training as well as during training with both responses rewarded, but these rats are likely to show such alternation on a given trial if and only if they have alternated on the previous trial.

Many theories of simple two-choice learning include the following assumption: The probability of choosing a particular response does not decrease when the response is chosen and rewarded, and does not increase when the response is chosen and not rewarded. This assumption, although generally consistent with the overall changes in behavior produced by various schedules of reward, is occasionally less successful in accounting for the detailed sequential properties of such behavior. Since most of these failures have been observed in rather specialized experiments, however, their importance is open to some question. Therefore, in order to determine more systematically the conditions under which the reward assumption is valid, it would be useful to derive from the assumption detailed consequences that can be tested in standard experimental situations. One such consequence has been stated by Bush (1963, pp. 466–467); this test has been used to demonstrate the influence of the gambler's fallacy in binary prediction experiments. Other consequences are derived in the present paper; these tests turn out to be particularly sensitive to spontaneous alternation in T-maze experiments.

Some response alternation is almost certain to occur in any learning process simply because of binomial variability and perhaps also because of changes in response probabilities produced by the reward schedule. Spontaneous alternation, on the other hand, is generally attributed to some additional cause such as fatigue, boredom, or curiosity. Clearly, it is unnecessary to invoke any such additional cause, however, unless simpler processes are proved to be insufficient. Hence, spontaneous alternation can be defined as alternation that cannot be explained in terms of the general reward assumption introduced in the previous paragraph. The present paper shows how much alternation is consistent with the assumption under various schedules of reward.

In the derivation of these results, the following notation will be used. Experimental trials will be numbered

---

[1] This paper is based on a dissertation submitted to the University of Pennsylvania in partial fulfillment of the requirements for the PhD degree. This work was supported by National Science Foundation Grants G-14839 and GB-1855, and by National Science Foundation Cooperative Predoctoral Fellowships. I would like to express my gratitude to Robert R. Bush for his advice and encouragement throughout the course of this research. I would also like to thank Elijah P. Lovejoy, Richard M. Rose, John van Laer, and David R. Williams for their helpful comments on the manuscript, Eugene Galanter for designing the apparatus, and George Croswell for helping to construct it.

[2] Now at Stanford University.

$1, 2, \ldots, n, \ldots$. Let r and r' denote the two responses available to the subject (S). Since some results are derived which apply to any particular sequence of experimental events such as stimuli, responses, or outcomes, let t denote an arbitrary sequence of such events. Let the indicator random variable X_n have the value 1 if the response r is chosen on Trial n, and the value 0 otherwise; similarly, let the random variable Y_n have the value 1 if the sequence of events t ends on Trial n, and the value 0 otherwise. Finally, let p_n represent the probability that S chooses response r on Trial n.

Two sequential properties of response data will be considered: the repetition of responses and the length of response runs. To represent repetition of either response, let $J_n = X_nX_{n+1} + (1 - X_n)(1 - X_{n+1})$; clearly, J_n has the value 1 if the same response is chosen on Trials n and $n + 1$, and the value 0 otherwise. To represent repetition of the response r following the sequence of events t, the product $Y_{n-1}X_nX_{n+1}$ can be used; this product has the value 1 if t ends on Trial $n - 1$ and r is chosen on Trials n and $n + 1$, and the value 0 otherwise. To represent short response runs, let $K_n = (1-X_{n-1})X_n(1-X_{n+1})$ and $K'_n = X_{n-1}(1 - X_n)X_{n+1}$; for instance, K_n has the value 1 if an r run lasting only one trial occurs on Trial n, and the value 0 otherwise. To represent longer runs, let $L_n = X_{n-1}X_n(1 - X_{n+1})$ and $L'_n = (1 - X_{n-1})(1 - X_n)X_{n+1}$; for instance, L_n has the value 1 if an r run lasting more than one trial ends on Trial n, and the value 0 otherwise. It should be noted that these run random variables do not count runs that are terminated by the end of an experiment or short runs that occur on the first trial of an experiment.

Clearly, the sum over trials of any of these random variables represents the total number of times the corresponding response repetition or response run occurs an experiment. To simplify notation, all such sums will be understood to run from $n = N'$ through $n = N - 1$. Predictions about the expected values of these sums will now be derived from various cases of the reward assumption. No predictions are made, however, about the variances of these sums, and it seems unlikely that such predictions can be derived without stronger assumptions. This limitation is not crucial, however, because variances can be estimated from data if necessary.

RESPONSE REPETITION

The hypothesis of the first theorem requires that a choice of either response increases the expected probability of choosing the same response again. According to the reward assumption, therefore, this theorem can be applied to experiments in which both responses are consistently rewarded.

Theorem 1

If for all p_n with $N' \leq n \leq N - 1$,

$$E(p_{n+1} | p_n, X_n = 1) \geq p_n$$

and

$$E(p_{n+1} | p_n, X_n = 0) \leq p_n,$$

then

$$\frac{E(\Sigma J_n)}{N - N'} \geq \tfrac{1}{2}.$$

Proof. For any p_n,

$$\begin{aligned} E(J_n | p_n) &= E[X_nX_{n+1} \\ &\quad + (1-X_n)(1-X_{n+1}) | p_n] \\ &= p_nE(p_{n+1} | p_n, X_n=1) \\ &\quad + (1-p_n)[1 \\ &\quad - E(p_{n+1} | p_n, X_n=0)] \\ &\geq p_n^2 + (1-p_n)^2 \\ &\geq \tfrac{1}{2}. \end{aligned}$$

The first five lines above are true by definition. The sixth line is derived by substitution of the hypotheses of the theorem in the third, fourth, and fifth lines. The last line follows because the function $p_n^2 + (1 - p_n)^2$ has an absolute minimum equal to $\frac{1}{2}$. Since this result holds for any p_n, it is possible to take expectations over all p_n; then, to complete the proof, it is necessary only to sum over n and reverse the order of summation and expectation.

The conclusion of Theorem 1 states that during Trials N' through $N - 1$, the expected number of repetitive responses should be at least half the total number of trials. In an experiment in which both responses are rewarded, therefore, an insufficiency of repetitive responses would suggest that the reward assumption is invalidated by spontaneous alternation. This theorem thus gives a rigorous justification of the conventional test for spontaneous alternation in experiments with reward for both responses. This test is not very sensitive, however, unless both responses are chosen about equally often: indeed, if one response is chosen on at least three-quarters of the trials, that response must necessarily be repeated on at least half the trials, whether or not the reward assumption holds. Moreover, the test is not even justified in experiments without reward for both responses; in the absence of additional information about the distribution of response probabilities and the form of the learning process, the reward assumption does not restrict the expected value of J_n on trials on which one or both responses are unrewarded.

Generalized Response Repetition

The hypothesis of the next theorem generalizes the hypothesis of Theorem 1 in two respects. First, for only one of the responses is performance required to increase expected response probability; thus, the theorem can be applied to experiments in which only one response is rewarded. Second, this increase in response probability is only required to occur if a particular sequence of events has ended on the previous trial; thus the theorem can be used to study the influence of previous events upon spontaneous alternation.

Theorem 2

If for all p_n with $N' \leq n \leq N - 1$,

$$E(p_{n+1} | p_n,\ Y_{n-1}X_n = 1) \geq p_n,$$

then

$$\frac{E(\Sigma Y_{n-1}X_nX_{n+1})}{E(\Sigma Y_{n-1}X_n)} \geq \frac{E(\Sigma Y_{n-1}X_n)}{E(\Sigma Y_{n-1})}.$$

Proof. By hypothesis, for any p_n,

$$\begin{aligned} E(X_nX_{n+1} | p_n,\ Y_{n-1} = 1) &= p_nE(p_{n+1} | p_n,\ Y_{n-1}X_n = 1) \\ &\geq p_n^2. \end{aligned}$$

Hence, if expectations are taken over all p_n, with Y_{n-1} equal to 1,

$$\begin{aligned} E(X_nX_{n+1} |\ Y_{n-1} = 1) &\geq E(p_n^2 |\ Y_{n-1} = 1) \\ &\geq [E(p_n |\ Y_{n-1} = 1)]^2. \end{aligned}$$

This result can be substituted in the numerator of the second fraction in the derivation below; the rest of the derivation follows from well-known properties of expectations and a few obvious cancellations:

$$\begin{aligned} \frac{E(Y_{n-1}X_nX_{n+1})}{E(Y_{n-1}X_n)} &= \frac{E(Y_{n-1})E(X_nX_{n+1} |\ Y_{n-1} = 1)}{E(Y_{n-1})E(X_n |\ Y_{n-1} = 1)} \\ &\geq \frac{[E(p_n |\ Y_{n-1} = 1)]^2}{E(p_n |\ Y_{n-1} = 1)} \\ &= E(X_n |\ Y_{n-1} = 1) \\ &= \frac{E(Y_{n-1}X_n)}{E(Y_{n-1})}. \end{aligned}$$

Consequently, by cross-multiplication,

$$\sqrt{E(Y_{n-1})E(Y_{n-1}X_nX_{n+1})} \geq E(Y_{n-1}X_n).$$

This result can be substituted to give the third line below; the first two lines follow from the Schwartz inequality:

$$\begin{aligned}[\Sigma E(Y_{n-1})]\cdot[\Sigma E(Y_{n-1}X_nX_{n+1})] &\geq [\Sigma\sqrt{E(Y_{n-1})E(Y_{n-1}X_nX_{n+1})}]^2 \\ &\geq [\Sigma E(Y_{n-1}X_n)]^2.\end{aligned}$$

To complete the proof, it is necessary only to cross-divide and reverse the order of summation and expectation.

This theorem states that during Trials N' through $N-1$, the expected number of times the event t is followed by at least two consecutive choices of the response r, divided by the expected number of times t is followed by at least one choice of r, must be at least as great as the expected number of times t is followed by at least one choice of r, divided by the expected number of times t occurs. Thus, Theorem 2, like Theorem 1, places a lower bound on the expected number of repetitive choices, but the lower bound in Theorem 2 depends upon the relative frequency of the responses. This theorem can therefore be used to detect spontaneous alternation even if one response is chosen much more often than the other.

A special case of Theorem 2 is of particular interest. If both responses are rewarded, the theorem can be applied to r' as well as to r; in order to eliminate the dependence upon previous events, t can be identified with some event that occurs on every trial. This case of the theorem leads to the following lower bound on the expected number of repetitive choices of either response:

$$\frac{E(\Sigma J_n)}{N-N'} \geq \left[\frac{E(\Sigma X_n)}{N-N'}\right]^2 + \left[1 - \frac{E(\Sigma X_n)}{N-N'}\right]^2.$$

This result thus justifies, in experiments with reward for both responses, the test suggested by Dember and Fowler (1958) for detecting spontaneous alternation in spite of response biases. The above result is also clearly stronger than the conclusion of Theorem 1. In other words, Theorem 2 contains previous tests as a special case.

RESPONSE RUNS

The hypothesis of the third theorem requires that the expected probability of choosing r increase on each trial, no matter which response was chosen on the previous trial. The specification of the previous response (that is, X_{n-1}) appears to be mathematically necessary but psychologically unimportant. According to the reward assumption, this theorem can be applied to experiments in which one response is rewarded and the other response is not rewarded.

Theorem 3

If for all p_n with $N' \leq n \leq N-1$,

$$E(p_{n+1}\,|\,p_n, X_{n-1}) \geq p_n$$

for X_{n-1} equal to 1 or 0, then

$$E(\Sigma L'_n - \Sigma K_n)\epsilon[0, 1]$$

and

$$E(\Sigma L_n - \Sigma K'_n)\epsilon[-1, 0],$$

where K_n, K'_n, L_n, and L'_n are the run random variables defined earlier.

Proof. By hypothesis, for any p_n and X_{n-1},

$$\begin{aligned}E(X_{n+1} - X_n\,|\,p_n, X_{n-1}) &= E(p_{n+1}\,|\,p_n, X_{n-1}) - p_n \\ &\geq 0.\end{aligned}$$

Hence, if expectations are taken over all p_n, for any X_{n-1},

$$E(X_{n+1} - X_n | X_{n-1}) \geq 0. \qquad [1]$$

According to the definitions of the run random variables,

$$\begin{aligned} L'_n - K_n &= (1 - X_{n-1})[(1 - X_n)X_{n+1} \\ &\qquad - X_n(1 - X_{n+1})] \\ &= (1 - X_{n-1})(X_{n+1} - X_n); \end{aligned}$$

$$\begin{aligned} K'_n - L_n &= X_{n-1}[(1 - X_n)X_{n+1} \\ &\qquad - X_n(1 - X_{n+1})] \\ &= X_{n-1}(X_{n+1} - X_n); \end{aligned}$$

$$L'_n - K_n + K'_n - L_n = (X_{n+1} - X_n).$$

It is next possible to take expectations in these results and then to substitute Equation 1:

$$\begin{aligned} &E(L'_n - K_n) \\ &\quad = E(1 - X_{n-1}) \\ &\qquad \cdot E(X_{n+1} - X_n | X_{n-1} = 0) \\ &\quad \geq 0; \qquad [2] \end{aligned}$$

$$\begin{aligned} &E(K'_n - L_n) \\ &\quad = E(X_{n-1}) \\ &\qquad \cdot E(X_{n+1} - X_n | X_{n-1} = 1) \\ &\quad \geq 0; \qquad [3] \end{aligned}$$

$$E(L'_n - K_n) + E(K'_n - L_n) = E(X_{n+1} - X_n).$$

Because the expectations in Equations 2 and 3 are both nonnegative, neither can exceed their sum. Hence,

$$E(L'_n - K_n) \leq E(X_{n+1} - X_n); \qquad [4]$$

$$E(K'_n - L_n) \leq E(X_{n+1} - X_n). \qquad [5]$$

To prove the first statement in the conclusion of the theorem, it now suffices to sum Equations 2 and 4 over n and reverse the order of summation and expectation:

$$E[\Sigma(L'_n - K_n)] \geq 0;$$

$$\begin{aligned} E[\Sigma(L'_n - K_n)] &\leq E[\Sigma(X_{n+1} - X_n)] \\ &= E(X_N - X'_N) \\ &\leq 1. \end{aligned}$$

The second statement in the conclusion can be derived from Equations 3 and 5 in a similar manner.

The conclusion of Theorem 3 states that during Trials N' through $N - 1$, the expected number of long runs of one response should not differ by more than 1 from the expected number of short runs of the other response. This conclusion implies, moreover, that the expected total number of long runs should not differ by more than 1 from the expected total number of short runs. In a simple position-training experiment, therefore, a predominance of short runs would indicate that the reward assumption is invalidated by spontaneous alternation, and a predominance of long runs would indicate that the reward assumption is invalidated by spontaneous perseveration. Thus, Theorem 3, unlike the previous theorems, can be used to study perseveration as well as alternation. For instance, this theorem could be used to determine whether the increases in perseverative behavior observed to follow certain brain lesions and drug treatments actually represent spontaneous perseveration not explainable by simple reward effects. At present, however, data suitable for this purpose do not seem to be available.

EXPERIMENTAL EVIDENCE

Most of the previous experiments on spontaneous alternation have already been reviewed by Dember and Fowler (1958). The criterion established in Theorem 1 has frequently been used as a test for spontaneous alternation in experiments in which both responses were rewarded and also, incorrectly, in experiments in which neither response was rewarded. Fortunately, the presence or absence of reward does not appear to have made much difference in these studies. The variable that seems most strongly to have influenced alternation is the distribution of trials. Heathers (1940), Montgomery (1951), Riley and Shapiro (1952), and Walker (1956) have all reported that spontaneous alternation occurs with massed trials, but not with spaced trials.

Theorems 2 and 3 permit the study of spontaneous alternation to be pursued in greater detail and extended to experiments without reward for both responses. To illustrate the use of these theorems, a new series of experiments was run, and a new analysis was performed on data previously reported by Galanter and Bush (1959) and Fey (1960, 1961). In the new experiments and also in the Galanter-Bush and Fey experiments, rats were run in a T maze with food reward. In some of the new experiments, the general conditions of previous studies were reproduced by giving reward for both responses with massed trials. In the rest of the new experiments, these conditions were changed by giving reward for only one response with massed trials. In the Galanter-Bush and Fey experiments, the conditions were further changed by giving reward for one response with spaced trials.

METHOD

The new experiments will now be described in greater detail. The subjects (*S*s) were 27 female Sherman-strain albino rats, obtained at the age of 90 days from Scott Farms, Stroudsburg, Pennsylvania. Of these rats, 4 were discarded early in training for failure to run the maze properly; the rest of this report concerns the remaining 23 *S*s.

The apparatus was an automatic T maze in which a rat could run for long sessions of consecutive trials without the aid of the experimenter (*E*). The stem section of this maze was 40 inches long, the arm sections were at right angles to the stem and 24 inches long, the goal sections were parallel to the stem and 24 inches long, and the sections through which the rat could return from the goal to the stem were at acute angles to the stem and 32 inches long. All sections were 4 inches wide and 5 inches deep. The walls were of ½-inch plywood, painted flat black, and the floor and ceiling were of hardware cloth. Sections were separated by guillotine doors; 12 inches beyond each door was a photobeam. When the rat passed through the photobeam, the door ahead was raised by a solenoid, and the door behind was dropped; the raising of the door could be delayed by a timer. A drinking tube projected into the stem section from a water bottle resting on top. A food cup was recessed into the wall of each goal section. When the rat passed through the photobeam in the goal section, a 45-milligram Noyes pellet could be dropped into the cup as a reward. The schedule of rewards was programmed in advance, and the sequence of choices made by *S* was recorded automatically.

The maze, control apparatus, and *S*s' individual home cages were all kept in the same room. The room was partially soundproof, continuously illuminated, and maintained thermostatically at 75° F. Water was always available to *S*s. Food, consisting of Purina lab chow and the Noyes pellets earned in the maze, was limited to 10 grams a day.

After 3–6 weeks of adaptation to handling and food deprivation, each *S* received a short program of manual pretraining. The rat was first given six Noyes pellets in its home cage, then placed in the goal section of the maze and rewarded, five times on each side, and finally placed in the arm section and rewarded in the goal section, five times on each side. The first experimental session immediately followed pretraining on the same day; later sessions were run at approximately the same time on consecutive days. The rat was placed in the stem section of the maze to start the first trial of a session and was removed from the return section after the last trial. If the rat failed to complete any trial within 5 minutes, the solenoids raising the doors were all automatically turned off to prevent over-heating; *E* entered the room and restored the power when the solenoids were cool. Such events occurred in less than one-sixth of the sessions; otherwise, *E* did not enter the room during a session.

Two groups of *S*s were given reward for both responses. In Group 1, *S*s were not delayed in any section of the maze; in Group 2, *S*s were delayed for 10 seconds in the stem section on each trial. Sessions contained 220 trials, because 220 pellets exactly completed a 10-gram ration. Groups 1 and 2 contained five and four rats, respectively. Except for two *S*s in Group 1 who were run for seven sessions, all *S*s were run for at least eight sessions. As the extra sessions did not reveal any changes in behavior, only the first eight sessions for each *S* will be considered henceforth.

Two other groups were given position and reversal training. In Group 3, *S*s were not delayed in any section; in Group 4, *S*s were delayed for 10 seconds in the stem section on each trial. Sessions contained 200 trials, divided into four 50-trial training periods. Reward was given on one side for the first training period, on the opposite side for the next period, and so on; half the *S*s were first rewarded on the right, and half on the left. Each group contained seven rats, all of which were run for four sessions.

The Galanter-Bush and Fey experiments have already been described in the original publications; the design of these experiments will be summarized here. Only the groups given conventional position and reversal training will be considered. In Groups 3 and 4 of the Galanter-Bush experiment, *S*s were given three trials a day, separated by 30 minutes, for a total of 192 trials. The 20 rats in Group 3 were rewarded on the right for the first 48-trial period, on the left for the next 48-trial period, and so on; the 19 rats in Group 4 were rewarded on the right for a 144-trial period, and on the left for a 48-trial period. In Group 1 of the Fey experiment, *S*s were given one trial a day for a total of 65 trials. The 25 rats in this group were rewarded on the right for a 30-trial period, and on the left for a 35-trial period.

RESULTS

Of *S*s given reward for both responses in the new experiments, nearly all developed fairly consistent biases toward one side or the other, although no *S* chose the same side on more than nine-tenths of the trials in any day. The overall percentages of choices of the left side for each *S* were as follows: in Group 1—30, 37, 50, 51, 71; in Group 2—26, 32, 41, 66.

Of *S*s given position and reversal training in the new experiments, all learned to choose the correct side on most trials toward the end of each 50-trial training period. Not surprisingly, Group 4, given a short delay between trials, appeared to learn slightly faster than Group 3. As in previous experiments on repeated reversal learning, both groups tended to learn faster in later training periods than in earlier periods. Compared to the groups given similar training in the Galanter-Bush and Fey experiments, Groups 3 and 4 appeared to learn more slowly in the first daily session, at about the same rate in the second session, and faster in the last two sessions.

The prediction stated in Theorem 2 can be tested in the new experiments and also in the Galanter-Bush and Fey experiments. According to the hypothesis of the theorem, the response *r* can be identified with either a left turn or a right turn in Groups 1 and 2, but must be identified with the correct response in the other groups; to investigate the influence of the previous response, the event *t* can be identified with either response. Each row of Table 1 presents tests of the prediction for a single group with a single identification of the response *r*. The Galanter-Bush and Fey experiments are denoted by obvious abbreviations, and the response identified with *r* is indicated in parentheses for Groups 1 and 2. In the tests labeled "Previous response same," *t* is identified with the same response as *r;* in the tests labeled "Previous response different," *t* is identified with the opposite response. Each test is a comparison of two ratios of sums;

TABLE 1

TESTS OF THE PREDICTION IN THEOREM 2

Group	Previous response same			Previous response different		
	Group tests[a]	*S* tests[b]	Block tests[b]	Group tests[a]	*S* tests[b]	Block tests[b]
1(R)	60 > 36	25–8	11–0	23 < 70	1–37	0–11
1(L)	57 > 30	22–7	11–0	19 < 64	0–38	0–11
2(R)	67 > 58	13–17	10–1	51 < 60	6–26	0–11
2(L)	62 > 40	16–12	11–0	26 < 42	1–31	0–11
3	92 > 75	28–0	8–1	48 < 71	5–23	1–10
4	91 > 85	27–1	9–0	58 < 60	12–16	1–10
GB3	90 > 84	20–0	8–1	62 > 53	14–6	7–4
GB4	95 > 92	19–0	4–1	62 > 35	19–0	5–0
F1	90 > 84	24–1	4–1	63 > 39	22–2	4–1

[a] Theorem 2 predicts that the second figure will not exceed the first.

[b] The first figure is the number of tests confirming the prediction, and the second figure is the number of tests contradicting the prediction.

three sets of tests, differing in range of summation, are presented. In the Group tests, the sums are taken over all *S*s and days for each group. The table states the ratios of these sums, expressed as percentages; the theorem predicts that the first percentage will be at least as large as the second. In the *S* tests, the sums are taken over each daily session for each *S* in the new experiments, and over all days for each *S* in the Galanter-Bush and Fey experiments. The table states the number of tests confirming and contradicting the prediction, not counting tests in which the denominator of one of the ratios is less than 5. In the Block tests, the sums are taken over all *S*s, and also over all days in the new experiments, for each of the following blocks of trials. In the groups given reward for both responses, a session is divided into 10 blocks of 20 trials followed by a block of 18 trials. In the groups given position and reversal training, the blocks are chosen to separate three different stages of training: early, when the group proportion of correct responses is less than about .35; middle, when the proportion correct is between .35 and .85; and late, when the proportion correct is greater than .85. Using these criteria, the first training period in a session is divided into a block of 20 trials, and then a block containing the rest of the period; except in Group 4 of the Galanter-Bush experiment, each later training period is divided into a block of 5 trials, then a block of 20 trials, and then the rest of the period; in the Galanter-Bush Group 4, the second period is divided into a block of 20 trials, then a block of 15 trials, and then the rest of the period. The table again states the numbers of tests confirming and contradicting the prediction, not counting tests in which the denominator of one of the ratios is less than 5.

The Group tests show a fairly clear pattern of results. In the Galanter-Bush and Fey experiments, the prediction is confirmed in all tests, but in the new experiments, the prediction is confirmed only when t and r are identified with the same response and is contradicted when t and r are identified with different responses. The *S* and Block tests generally although not unanimously corroborate these results. Most of the exceptions occur in the *S* tests: in Group 2 with

t the same as r, and in Group 4 with t different from r.

The prediction stated in Theorem 3 can be tested in all the groups given position and reversal training. The response r is always identified with the correct response. Each row of Table 2 presents tests of this prediction for a single group. Each test is a comparison of the mean number of long runs (of length greater than 1) per training period with the mean number of short runs (of length 1) per training period; two sets of tests, differing in range of averaging, are presented. In the Group tests, the means are taken over all *S*s and days for each group. The table states the mean number of long runs followed by the mean number of short runs; the means are indicated as equal if they do not differ by more than 1, as predicted in the theorem. In the *S* tests, the means are taken over each daily session for each *S* in the new experiments and over all days for each *S* in the Galanter-Bush and Fey experiments. The table states the numbers of tests showing more long runs than predicted, run lengths consistent with the prediction, and more short runs than predicted. Block tests are not presented because such tests are rendered relatively insensitive by the small numbers of runs that occur during short blocks of trials.

TABLE 2

TESTS OF THE PREDICTION IN THEOREM 3

Group	Group tests[a]	*S* tests[b]
3	5.7 < 12.2	0– 3–25
4	4.9 < 6.9	1– 7–20
GB3	5.4 = 6.2	3– 7–10
GB4	6.1 = 6.0	6– 7– 6
F1	3.9 = 3.4	9–12– 4

[a] Theorem 3 predicts that the two figures will be indicated as equal.

[b] The first figure is the number of tests showing more long runs than predicted, the second figure is the number of tests confirming the prediction, and the third figure is the number of tests showing more short runs than predicted.

The Group tests again follow a fairly clear pattern. In the Galanter-Bush and Fey experiments, the prediction is confirmed as before, but in the new experiments, too many short runs are observed. The *S* tests again generally corroborate these results. As before, more exceptions occur in Group 4 than in Group 3.

DISCUSSION

The foregoing analysis indicates that spontaneous alternation is more prevalent than has previously been demonstrated. Using Theorems 2 and 3, spontaneous alternation was detected in the new experiments although not in the Galanter-Bush or Fey experiments. Hence, spontaneous alternation requires massed trials but not reward for both responses. In other words, given massed trials, *S*s show spontaneous alternation not only when they have no reason not to alternate, but also when they must learn to repeat a correct response. This fact thus suggests that the original reward assumption is generally inapplicable in massed-trial T-maze experiments, and that such experiments are consequently not very useful for the detailed study of simple reward effects.

The more detailed tests also indicate that spontaneous alternation may depend upon previous responses. Using Theorem 2, spontaneous alternation was generally detected only when t and r were identified with different responses. Hence, *S*s are likely to show spontaneous alternation on a given trial if, and only if, they have alternated on the previous trial. Unfortunately, this finding could be an artifact caused simply by fluctuations in the overall frequency of alternation. Although such a possibility cannot in principle be completely eliminated, the *S* and Block tests suggest at least that the observed results are not attributable to differences between *S*s, days, parts of a day, or stages of training. It therefore seems reasonable to conclude that spontaneous alternation reduces the probability of a repetitive response at the beginning of a response run, but generally does not further reduce this probability if the run persists. In position and reversal training, therefore, such alternation may be expected to retard but not to prevent learning to repeat the correct response. This fact thus explains why the original reward assumption may be plausible in the long run although incorrect in detail.

REFERENCES

BUSH, R. R. Estimation and evaluation. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Vol. 1. New York: Wiley, 1963. Pp. 429–469.

DEMBER, W. N., & FOWLER, H. Spontaneous alternation behavior. *Psychological Bulletin*, 1958, **55**, 412–428.

FEY, C. Parameter invariance in models for learning. Unpublished doctoral dissertation, University of Pennsylvania, 1960.

FEY, C. An investigation of some mathematical models for learning. *Journal of Experimental Psychology*, 1961, **61**, 455–461.

GALANTER, E., & BUSH, R. R. Some T-maze experiments. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford University Press, 1959. Pp. 265–289.

HEATHERS, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *Journal of Psychology*, 1940, **10**, 359–380.

MONTGOMERY, K. C. "Spontaneous alternation" as a function of time between trials and amount of work. *Journal of Experimental Psychology*, 1951, **42**, 82–93.

RILEY, D. A., & SHAPIRO, A. M. Alternation behavior as a function of effortfulness of task and distribution of trials. *Journal of Comparative and Physiological Psychology*, 1952, **45**, 468–475.

WALKER, E. L. The duration and course of the reaction decrement and the influence of reward. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 167–176.

(Received June 30, 1965)

Psychological Review
1966, Vol. 73, No. 5, 437-444

PROCESS OF ENUMERATION[1]

MARY BECKWITH AND FRANK RESTLE

Indiana University

Mathematically, the process of enumeration is fundamental to arithmetic. Psychologically, it is a sensorimotor chain controlled at every stage by a shifting perceptual organization. Enumeration requires a chant ("1, 2, 3 . . ."), a shifting indicator response (pointing), and a perceptual grouping of objects into those already counted and those still ahead. The arrangement of the objects has, theoretically, an important effect on the speed and accuracy of enumeration. Further analysis shows that the serial chain of behavior, required for counting a fairly large set of objects, must be divided into parts, and the objects grouped into corresponding subsets. 3 experiments show the relationship between arrangement of objects and counting.

At the foundation of applied arithmetic lies the ability to enumerate sets. A basic procedure in counting is to chant the numerals, "ONE, TWO, THREE . . . ," at each word, pointing to one of the objects in the set to be counted. The chain of indicator responses is often generated from a basic response like pointing, by a modification like moving to the right. Then, if I is the indicator response and T is its transformation to get to the next object, the responses are I, T(I), T[T(I)], . . . , $T^n(I)$. The indicator responses must touch every object once and stop when all the objects have been counted. This requires perceptual control, a sliding discrimination between the set counted (C) and the set as yet uncounted (U). As objects are shifted from C to U it is seen that at any stage n, $C_{n+1} \supset C_n$, whereas $U_{n+1} \subset U_n$. The sequence of sets constitutes a *linear array* (Restle, 1959b, 1961), is ordered, and provides a basis for simple measurement of distance. Finally, the end of the process of enumeration must be recognized, usually by the emptiness of U. We shall refer to the above process as "simple enumeration."

Young children may fail to produce the chant correctly, but older subjects (*S*s) presumably have difficulty mainly in the sensorimotor components of counting. Some of the factors in sensorimotor control of counting are seen when the experimenter (*E*) varies the arrangement of the objects to be counted. In simple enumeration, if the objects lie in a straight line, then the transformation of the pointing response T is merely a move to (say) the right, and the discrimination between C and U can be made relative to the location of the finger. If the objects are in a circle, the transformation may again be a simple shift in the clockwise direction, but the discrimination between C and U becomes more difficult as *S* approaches the origin of his counting. This may lead to errors and perhaps to long delays if *S* starts over. When the objects to be counted are all alike, the circular arrangement should be very difficult, though the straight line is not affected. If the objects differ among themselves, say in shape, then the circular arrange-

---

[1] This research was supported in part by NSF Grant GB 2848 from the National Science Foundation to the second author. The authors wish to express their appreciation to the University School, Bloomington, Indiana, and to Richard C. Pell, principal, for his cooperation.

ment should become easier. With a rectangular arrangement, again there is a simple path, and there should be no difficulty with the stopping rule. In simple enumeration the rectangular arrangement has no great advantages over the straight line, and, since it provides some optional paths, may even introduce delays. Finally, the objects may be scrambled—approximately equally spaced but in no particular pattern. This forces *S* to devise a path through the set; he must use a complex sequence of transformations of the indicator responses, and, since no single cue distinguishes elements of C from U, *S* must construct and remember cues as he goes along.

Simple enumeration is not a complete psychological description of the process of enumeration because it takes no account of the fact that assemblies of discrete objects are perceived in groups. Of Wertheimer's six principles (Woodworth & Schlosberg, 1954, p. 409) the authors' experiments permit variation of proximity, similarity, and good continuation. The *S*s presumably perceive the field of objects as organized into groups. A natural question is whether these groupings have any function in counting.

Enumerating as many as 16 objects is a long serial task, and *S* may organize the segments into a convenient number of smaller subtasks. If so, then it is necessary to group objects conveniently to permit rapid and accurate counting; for example, if *S*s need to group objects, the rectangular arrangement may be consistently better than the straight line.

HISTORY OF THE PROBLEM

Although counting is basic to applied arithmetic, there is little literature of experimental studies dealing with counting. Many studies have dealt with the immediate apprehension of number—a process supposed not to be mediated by steps, but to occur instantly. Freeman (1912) set the span of apprehension of circles of light at 5 for adults and 4 for children over 8 years; younger children were untestable. Fernberger (1921) used random arrangements of 4–12 dots and found that the relative frequencies of correct responses made a continuous ogival function of number of dots shown. Correct judgments were made 50% of the time (Fernberger's measure of visual discrimination) in the range from 6 to 11 dots. Hunter and Sigler (1950) varied intensity and duration of presentation of dots and found that the usual visual laws applied.

One approach to the span-of-apprehension idea is as detection of a number of targets (Casperson & Schlosberg, 1950; Restle, Rae, & Kiesler, 1961; Schlosberg, 1948). This approach deals mainly with perception of a number of dots under short exposures or poor illumination, but may undermine the concept of a "magical number."

Warren (1897) arranged white squares in a vertical line on a black background, measured the time taken to count them, and concluded that 4 was the limit of perceptual counting. Bourdon (1908) also used the method of reaction time and found a small increase in time as a function of number of dots to be counted. He judged that 7 was the limit of the number that could be accurately ascertained at one glance. Saltzman and Garner (1948) used concentric rings as targets and decided that the method of reaction time was useful for the study of perception of number.

Von Szeliski (1924) found that fields of up to 6 dots or figures could be perceived without eye movement. Jensen, Reese, and Reese (1950) used fields of up to 29 dots and found that errors appeared first in 11-dot fields and became more frequent in larger fields. After 6 dots, the plot of reaction time against number of dots becomes more scattered, and a more shallow, positively accelerated curve appears. From this, Jensen et al. concluded that *S*s subitize up to 5 or 6 dots.

Taves (1941) studied judgments of numerousness (considered a stimulus variable like loudness) and stated that two mechanisms were involved—one used for up to 7 dots, the other for larger fields. He found that judged numerousness was decreased if the dots were arranged into a single figure, for example, a circle. Kaufman, Lord, Reese, and Volkmann (1949) also suggested two separate discriminatory mechanisms—one acting for up to 6 dots, the other for larger collections.

All of these studies concentrate upon rapid apprehension or judgment of number and are cast in a theoretical framework drawn from perception. Much effort has gone into determining the number of things that can be apprehended at a single glance, even though neither civilized adults nor civilized children count at a glance.

In summary, the two main methods of studying enumeration are the tachistoscopic, in which a display is briefly displayed and judged, and the reaction-time experiment, in which the material is displayed continuously, but *S* responds as quickly as possible. Both methods are perceptual in character, though the reaction-time method permits some more complex methods of enumeration to occur. The main question has been the maximum number of dots that can be perceived at a glance. This number has been estimated from about 4 to 7 and clearly depends upon *S*s and the display used.

It is amazing that although everyone knows that objects are enumerated by counting, most studies of enumeration or the judgment of number have attempted to rule out counting and ensure that only the primitive method of guessing is employed. Classical psychological theory assumes that perception is a basic and simple process from which other, more complex cognitions are to be derived; furthermore, the task of experimental psychology is to analyze compound experiences into their elements. Following these tenets, the process of counting is reduced to its elements, "perception" of numbers, and no effort is expended in synthesizing the process of counting.

The experiments reported below have a different concept of enumeration. First, it is assumed that enumeration is a compound task, but that the true and interesting structure of the task is not found by analysis into bits, but by determining how the parts go together. Second, it is assumed that perception of small numbers may be a skill developed by adults, a sort of shortcut to counting, rather than an elementary mental event. Throughout, the authors will use the reaction-time method. The *S* is instructed to enumerate as quickly as possible.

EXPERIMENTS I AND II

Method

Subject. The first of these experiments used 18 students in an ungraded class in the summer of 1964, 10 boys and 8 girls, ranging in age from 7 years, 3 months to 9 years, 10 months. The second experiment used 20 college students, ages 17–22, drawn from the elementary psychology course at Indiana University.

Apparatus and procedure. The stimulus cards were 8½ × 11-inch sheets of white paper on which objects to be counted were outlined in black ink. Each *S* was run separately; each sat across a table from *E*, in an otherwise unused classroom which was arranged to reduce external visual distraction.

Each card was presented flat on the table, first shielded with an aluminum panel on a handle which was raised at each trial to display the card while a standard timer was engaged simultaneously. Response and latency of response were recorded. Time was spent gaining the attention and cooperation of each child whereas somewhat more formal instructions were given to the college students.

Experimental design. Each *S* counted the objects on 40 stimulus cards which were presented in a different random order for each *S*. Eight of these were fillers inserted to make the repeated appearance of certain numbers less obvious.

Variables were (*a*) number of objects to be counted, (*b*) arrangement of objects—line, circle, rectangle, scrambled, (*c*) all objects the same shape or different shapes.

Because of the rectangular arrangements, numbers with two fairly equal factors were employed—12, 15, 16, and 18. Filler cards had 5, 9, 10, 11, 13, 14, or 17 objects. Objects were spaced about ½ inch apart, with the straight line being horizontal, and the rectangles being 4 × 3, 5 × 3, 4 × 4, and 6 × 3. Objects on a card were either all the same shape or divided equally among squares, circles, isosceles triangles, and teardrop shapes. The circle's diameter was ½ inch; all other figures were of approximately equal apparent size. Thus, 4 Numbers × 4 Shapes × Same or Different yields 32 displays in a simple factorial design, replicated once per *S*.

Results

Most children employed pointing and chanting, though some counted silently and sometimes without pointing, while college students counted silently, occasionally pointing. All children counted all displays, making 106 errors out of 576 counts (18.4% error) with a mean time of .60 second per object. The college students averaged .30 second per object and made only 11.6% errors.

Larger sets of objects took longer to count for both groups. Linear equations fit both sets of data well, being $\bar{t} = .75(N - 3)$ for the children and $\bar{t} = .35(N - 2)$ for the college students, where $\bar{t}$ is a mean reaction time in seconds, and N is the number of objects to be counted. Since no measurements were taken with small numbers of objects, there is no reason to believe the straight line may be extrapolated back without curving (see Jensen et al., 1950).

Object arrangement influences both groups' response latencies, with the children counting the Rectangle most rapidly, then the Line, the Circle, and the Scrambled arrangement, in that order. The proportions of errors for the children were: Rectangle, .097; Line, .104; Circle, .215; and Scrambled, .326, thus showing the Scrambled arrangement most difficult both in time and errors. For college students, there was little difference between Line, Circle, and Scrambled, but the Rectangle was much faster.

Discussion

The results seem to show a difference between children's and college students' use of spatial arrangement. Children may use spatial arrangement to get a simple ordering of the stimuli, while college students seem to make special use of the rectangular array, presumably by using multiplication. The children may not multiply faster than they enumerate or they may not have learned to apply the concept of multiplication to this situation.

When they cannot multiply, college students and perhaps children may divide the set of objects, possibly by perceptual principles of grouping, into smaller subsets, enumerate the subsets, then add the numbers. Sets of up to 5 or 6 may be subitized by skilled adult *S*s. Calculation speed should depend upon convenience of grouping, enumeration, and addition. Rectangular arrays enforce a convenient subdivision which permits several simultaneous enumerations. Indeed, both groups consistently count the rectangular array rapidly, but for the college students the array of 16 objects is especially

TABLE 1

MEAN TIME TO COUNT (IN SEC.) BY YOUNG CHILDREN

No. objects	Arrangement of objects															
	Line				Circle				Rectangle				Scrambled			
	Same	Error	Diff.	Error	Same	Error	Diff.	Error	Same	Error	Diff.	Error	Same	Error	Diff.	Error
12	6.58	1	6.36	1	8.05	5	6.84	1	6.10	1	5.62	1	8.77	4	7.36	3
15	9.28	2	8.45	0	9.45	3	9.13	3	7.75	2	8.32	0	9.98	7	11.02	6
16	9.11	3	8.63	3	9.04	5	10.51	4	8.13	2	8.68	3	10.49	6	11.42	8
18	11.14	3	8.94	2	10.94	4	12.38	6	10.47	4	9.52	1	12.36	7	13.79	6

Note.—Proportion of errors is as follows: Line, .104; Circle, .215; Rectangle, .097; Scrambled, .326.

fast. This array was 4 × 4, and previous informal observations have shown that older children, and presumably college students, are faster at squaring than at other multiplications. Sets of 16 are not counted unusually rapidly in other arrangements, or in any arrangement by young children, so there is reason to believe that the advantage lies in multiplication.

It may also be noticed that rectangular arrays are counted more rapidly if all the objects are the same, whereas with circular and linear arrangements counting is faster with different objects. Rectangular arrangement enforces a clear grouping on the set; differentiation of figures only competes with this grouping. Spatial grouping is more difficult with straight lines and circles, and *S*s may use dissimilarities among figures to provide the necessary groupings.

EXPERIMENT III

In the first two experiments, only position and shape of the objects were varied. In the third experiment, color variation is also used so as to establish alternative perceptual groupings.

The general design of this experiment is based upon earlier studies of additivity of cues in cue learning (Bourne & Restle, 1959; Bower & Trabasso, 1963; Restle, 1955, 1959a, 1962; Trabasso, 1960, 1963). In such experiments several problems are constructed using two or more dimensions. One group may learn to classify drawings based on their shape, disregarding color. Another group may classify drawings using color, disregarding shape. A third group learns a classification in which color and shape are redundant, which is easier than the other two (additivity of cues).

TABLE 2

MEAN TIME TO COUNT (IN SEC.) BY COLLEGE STUDENTS

No. objects	Arrangement of objects															
	Line				Circle				Rectangle				Scrambled			
	Same	Error	Diff.	Error	Same	Error	Diff.	Error	Same	Error	Diff.	Error	Same	Error	Diff.	Error
12	3.92	0	3.48	0	4.24	2	4.23	2	1.92	1	2.15	2	3.57	1	3.55	2
15	5.28	4	4.83	3	5.51	4	4.70	3	3.20	0	2.67	0	4.59	2	4.67	2
16	5.59	4	4.97	2	5.99	1	5.83	1	2.19	1	2.70	5	5.45	5	5.45	4
18	6.07	2	5.63	0	6.53	3	6.31	7	3.34	2	3.81	0	5.54	0	6.13	5

TABLE 3

MEAN TIME TO COUNT (IN SEC.), EXPERIMENT III

No. objects	Conditions									
	Color and shape		Color		Shape		Ambiguous		Random	
	Time	Errors	Time	Errors	Time	Errors	Errors	Errors	Time	Errors
12	3.18	1	3.17	1	3.56	3	4.39	2	3.80	3
15	4.39	1	4.52	5	5.37	2	4.76	5	5.16	4
16	4.55	0	4.32	1	5.56	4	4.47	2	5.21	0
17	4.82	9	5.44	1	6.93	5	5.83	2	6.72	5
18	5.39	7	5.96	4	6.40	6	5.93	1	7.12	6

Several mathematical models of cue learning have given quantitative predictions of learning speed based on the number of redundant relevant cues, the number of irrelevant cues, and the weights and strengths of such cues.

Method

Subjects. Twenty-four college students, aged between 17 and 20, from an introductory psychology class served as *S*s.

Apparatus and procedure. These were identical with Experiment II.

Materials and design. Thirty displays, each consisting of a collection of shapes cut from construction paper and pasted to 8½ × 11-inch sheets of white paper were used. The shapes were four colors: brick red, Air Force blue, light blue-green, and light violet. They were four shapes: circle, square, triangle, and parallelogram. Test items had 12, 15, 16, 17, and 18 figures, and 5 filler presentations. In Condition 1, Color and Shape, the four colors and shapes were segregated, so that, for example, one section of the figure was all blue squares, another part red circles, etc. In Condition 2, Color, the four colors were segregated, but the shapes were scattered randomly over the field. In Condition 3, Shape, the four shapes were segregated, but the colors were scattered randomly over the field. In Condition 4, Ambiguous, both the four colors and the four shapes were segregated, but differently, so that the four circles might be half blue and half red, and of the three triangles, two might be red, and one violet, etc. In Condition 5, Random, both color and shape were scattered independently and randomly over the field.

Hypothesis. If rapid counting in the college student depends upon grouping the field of objects, then Condition 1, having a strong perceptual organization based on color and shape, should be counted most rapidly. Conditions 2 and 3, having color and shape arrangements, respectively, should provide estimates of the relative grouping value of the two dimensions. Condition 4 gives *S* two conflicting modes of organization. If one mode is much stronger than the other, the condition should be about as fast as the condition employing only the strong variable. Nearly equal variables might lead to instability of grouping, perhaps forcing *S* to start over in his counting. Condition 5, having no principle of grouping available, should lead to counting rates close to those observed in Experiment II, scrambled arrangements for college students.

Results

Time to count was a linear function of total number of objects. The results here are very close to those obtained with college students for the Scrambled condition in Experiment II. This indicates that the general findings are probably quite stable with successive samples from this pool of *S*s.

As predicted, the Color and Shape condition is counted fastest. Condition 2, Color, is almost as fast, and Condition 3, Shape, is slower, so it appears that color is a much more effective dimension than shape in this experiment. This agrees with many experimental studies comparing the two variables, varied approximately as in this experiment.

In these data, shape segregation has little advantage over Random, Condition 5. Therefore, it is predictable that the Ambiguous condition, having one strong (color) and one very weak (shape) organizing variable in competition, would be almost as fast as color alone. Notice that it is faster than Shape (Condition 3) or Random (Condition 5), though a small effect of shape is found in the fact that Condition 4 is slower than Conditions 1 or 2.

An analysis of variance showed that both the number of objects and the five conditions produce significant ($p < .01$) effects on speed of counting. These results establish that rapid counting depends upon grouping the material into subgroups. It is probable that *S*s segregated the whole display into color groups, subitized the number in each group, and then added these numbers to obtain the result. When no such grouping was possible, *S* was forced to rely on a relatively difficult and ambiguous grouping by position, or to count one by one.

DISCUSSION

The experiments reported above show that in counting, the set of objects is grouped perceptually, in accord with Wertheimer's rules of propinquity, good continuation, and similarity. This grouping then serves as a support or basis for counting, in that *S* performs his counting operation within groups and then in some fashion connects the various partial counts. One hypothesis is that college students *subitize* within each group, that is, determine the number by a somewhat mysterious but very rapid and accurate "perceptual" method. The *S* might collect such numbers and add them at the end or might subitize each group and add the result to a running total.

The fact that fairly young children show great sensitivity to the organization of the visual field, as shown in Experiment I, suggests that grouping may play a more general role in counting. That is, even when a child is enumerating one by one, he may work rapidly and routinely within one group, pause and consolidate or "store" his result in some way, and then attack the next group. The pausing, and the ability to divide the task into suitable parts, is possibly a generally important part of a long serial task.

REFERENCES

BOURDON, B. Sur le temps nécessaire pour nominer les nombres. *Revue Philosophique,* 1908, **65**, 426–431.

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychological Review,* 1959, **66**, 278–296.

BOWER, G. H., & TRABASSO, T. R. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford University Press, 1963. Pp. 32–94.

CASPERSON, R. C., & SCHLOSBERG, H. Monocular and binocular intensity thresholds for fields containing 1–7 dots. *Journal of Experimental Psychology,* 1950, **40**, 81–92.

FERNBERGER, S. W. A preliminary study of the range of visual apprehension. *American Journal of Psychology,* 1921, **32**, 121–233.

FREEMAN, F. N. Grouped objects as a concrete basis for the number idea. *Elementary School Teacher,* 1912, **12**, 306–314.

HUNTER, W. S., & SIGLER, M. The span of visual discrimination as a function of the time and intensity of stimulation. *Journal of Experimental Psychology,* 1950, **26**, 160–179.

JENSEN, E. M., REESE, E. P., & REESE, T. W. The subitizing and counting of visually presented fields of dots. *Journal of Psychology,* 1950, **30**, 363–392.

KAUFMAN, E. L., LORD, M. W., REESE, T. W., & VOLKMANN, J. The discrimination of visual number. *American Journal of Psychology,* 1949, **62**, 498–525.

RESTLE, F. A theory of discrimination learning. *Psychological Review,* 1955, **62**, 11–19.

RESTLE, F. Additivity of cues and transfer in discrimination of consonant clusters. *Journal of Experimental Psychology,* 1959, **57**, 9–14. (a)

RESTLE, F. A metric and an ordering on sets. *Psychometrika,* 1959, **24**, 207–220.

RESTLE, F. *Psychology of judgment and choice.* New York: Wiley, 1961.

RESTLE, F. The selection of strategies in cue learning. *Psychological Review,* 1962, **69**, 329–343.

RESTLE, F., RAE, J., & KIESLER, C. The probability of detecting small numbers of dots. *Journal of Experimental Psychology,* 1961, **61**, 218–221.

SALTZMAN, I. J., & GARNER, W. R. Reaction time as a measure of span of attention. *Journal of Psychology,* 1948, **25**, 227–241.

SCHLOSBERG, H. A probability formulation of the Hunter-Sigler effect. *Journal of Experimental Psychology,* 1948, **38**, 155–167.

TAVES, E. H. Two mechanisms for the perception of visual numerousness. *Archives of Psychology,* 1941, No. 265.

TRABASSO, T. R. Additivity of cues in discrimination learning of letter patterns. *Journal of Experimental Psychology,* 1960, **60**, 83–88.

TRABASSO, T. R. Stimulus emphasis and all-or-none learning in concept identification. *Journal of Experimental Psychology,* 1963, **65**, 398–406.

VON SZELISKI, V. Relation between quantity perceived and the time of perception. *Journal of Experimental Psychology,* 1924, **7**, 135–147.

WARREN, H. S. The reaction time of counting. *Psychological Review,* 1897, **4**, 569–591.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* (Rev. ed.) New York: Holt, 1954.

(Received July 6, 1965)

Psychological Review
1966, Vol. 73, No. 5, 445–458

COGNITIVE AND DRIVE FACTORS IN THE EXTINCTION OF THE CONDITIONED EYE BLINK IN HUMAN SUBJECTS[1]

KENNETH W. SPENCE

University of Texas

Extinction of the eyelid conditioned response (CR) in humans has been shown to be a function of factors relating to Ss' capabilities to discriminate the change in stimulus conditions at the inception of extinction. Hypothesizing that the rapid extinction obtained with humans in the standard conditioning situation results from an inhibitory set not to respond based on this discrimination, subsequent experiments in a masking situation designed to control (eliminate) this cognitive factor revealed that extinction of the eyelid response in humans proceeds at a relatively slow rate, as in animals. The findings with the masking situation also suggest that the partial reinforcement effect is absent in humans as in the case of classical aversive extinction with animals.

The *empirical construct* type of theory espoused by Hull (cf. 1943, 1952) and Spence (1956, 1958, 1960) takes its start from the findings of laboratory studies of classical and instrumental conditioning. That is, in this type of theorizing the specification of the quantitative properties of the theoretical concepts is based on empirical laws found in these simple kinds of experiments. Progress in the development of this theory has been most rapid in the areas of acquisition and motivation. On the basis of laws relating various response measures to such experimental variables as the number of reinforcements, magnitude of the reinforcing event, and deprivation operations, theoretical constructs such as associative or habit strength (H), incentive motivation (K), and drive level (D) have been introduced.

THEORETICAL INTERPRETATIONS OF EFFECTS OF NONREINFORCEMENT: CONCEPTS OF INHIBITION

Much less satisfactory is the portion of the theory that deals with the role of the nonreinforcement variable in determining response strength. In his *Principles of Behavior* Hull (1943) introduced the constructs of work inhibition (I_R) and conditioned inhibition (S^IR) to handle this area of behavior. While these constructs might have some validity in relation to the phenomenon of response effort, they have proved entirely inadequate as far as the behavioral phenomena resulting from nonreinforcement are concerned.

Actually, there was relatively little data on extinction phenomena available to Hull at the time he formulated his theory. But what little there was had led the present author to a very different conception from the one Hull (1943) proposed. As early as 1936, in a theoretical article on discrimination learning, the present author interpreted nonreinforcement of an appetitionally motivated response as involving frustration (Spence, 1936). While time limitations did not permit a discussion of extinction phenomena in the Silliman Lectures (Spence, 1956), in connection with the treatment of delay of reinforcement, it was pointed out that the two experimental operations, delay of reinforcement and nonreinforcement, were highly similar and presumably produced the same kind of behavioral effects. It was further suggested that underlying these effects were frustration-aroused competing responses. Since these Lectures, a number of my former students and I have further elaborated this frustration theory of inhibition, particularly in relation to a fairly sizable body of quantitative phenomena that have been shown to occur in instrumental reward conditioning and extinction under different schedules of reinforcement (cf. Amsel, 1958; Spence, 1960).

---

[1] This paper is based on colloquium talks given at a number of universities. Preparation for publication was supported by Contract NOnr 375(18) between the Office of Naval Research and the University of Texas.

Contrasting sharply with these recent developments in the instrumental reward type of conditioning, there has been little or no advance in our knowledge concerning experimental extinction in classical reward conditioning of the Pavlovian type. Unfortunately, the Russians did not go in for the kind of parametric studies that provide the type of quantitative laws that are needed for such theorizing as ours. As a consequence, it is not known to what extent the relevant variables and laws of extinction in classical reward conditioning are similar to or different from those discovered in the instrumental variety. Needless to say, this deficiency precludes the possibility of ascertaining whether our theory, as inferred from instrumental data, is sufficiently general to encompass classical phenomena. It is the author's own expectation that the laws of extinction of the classically rewarded response will be sufficiently different from those of instrumental reward learning as to require a somewhat different theoretical formulation. For example, the author does not believe that frustration-aroused, incompatible responses play the kind of role in the extinction of the classical response that has been conceived as occurring in the extinction of the instrumental response. But until more precise empirical data have been obtained concerning the effects of such variables as the magnitude and schedule of the original reinforcing stimulus on the subsequent extinction rate of the classical salivary response, such speculation is not very fruitful. There simply are too many alternative theoretical possibilities to choose among, and, since the experiments have to be conducted if the laws are to be obtained, we had best get to the job of doing them.

What has just been said about classical reward conditioning also holds for aversive types of conditioning. Certainly failure of occurrence of the noxious unconditioned stimulus (UCS) such as occurs in the extinction of a classical defense response can hardly be conceived as frustrating to the organism. Obviously, some other theoretical interpretation must also be found for this type of situation, other, that is, than that developed for instrumental reward conditioning. But again, it should be emphasized that the pressing need is for empirical studies that will provide the laws describing the course of experimental extinction of such responses and the variables that influence it. The remainder of this paper will be concerned with the findings of such studies, most of which were conducted at the University of Iowa under contract with the Office of Naval Research.

PROBLEM OF HIGHER PSYCHOLOGICAL PROCESSES IN HUMAN CONDITIONING STUDIES

The studies of classical defense conditioning conducted in our laboratory have involved the use of human subjects (college students), and the response conditioned has been the eye blink or corneal reflex as elicited by a puff of air directed at the eye. A major problem that arises in such research is that of controlling, with a view toward eliminating, the influence of higher psychological (symbolic) processes on these simple learned responses. To the extent that such higher activities do occur in these experiments the laws obtained are a function of unknown and unwanted variables. Certainly, such uncontrolled experiments would not provide the laws of *simple* learning that are being sought and thus would be of little value for the present theoretical endeavors.

Acquisition Phase

The author's evaluation of these studies is that, at least as far as the acquisition phase of conditioning experiments is concerned, the data (i.e., laws) they have provided are fairly adequate—in the sense that they are more or less free from the influence of these complicating factors. However, the same cannot be said for the data on the extinction phase. Indeed, it is strongly suspected that complex processes, call them cognitive factors, sets, mediating processes, or what you will, have greatly affected the extinction data heretofore reported for human subjects. If such is the case, then obviously these data do not provide a satisfactory basis on which to construct that part of the present theory concerned with the effects of nonreinforcement.

There is not sufficient time to document the basis of the author's optimism concerning acquisition data. It is enough to note that, on the whole, quite consistent data on acquisition are obtained from one study to another and across different laboratories. But more important, the laws relating the conditioning performance of humans to the several variables that have been investigated appear to have the same form as those obtained with animals. The one major difference—one, incidentally, that does not seem to be too well known—is that classical conditioning occurs with much greater rapidity in humans than animals. While this finding suggests different rate-of-conditioning parameters for different species, it is also possible that it merely reflects differences in the motivational level of different species under the same intensity of the UCS. For example, the extremely slow conditioning of the eyelid response in the rabbit may be due to a lesser sensitivity of the rabbit's cornea. It is known that the rate of conditioning in human subjects does vary with the intensity of the UCS. More research is needed with animal subjects under different UCS intensities before a correct interpretation can be made. The author would merely like to add that as a consequence of the early studies of Hilgard and his students on the effects of different kinds of verbal instructions, it appears that it is known how to produce "sets" in human subjects which minimize the influence of these cognitive factors, at least during the acquisition phase of conditioning (cf. Hilgard, Campbell, & Sears, 1938; Grant, 1939).

Extinction Phase

Turning now to the extinction data reported heretofore, at least two major discrepancies between the findings obtained with human and animal subjects are found, both of which led the author to suspect that the extinction performance of the humans was greatly influenced by cognitive factors not present in animals. The most striking of these differences and the one with which this paper will be primarily concerned is that in animals the extinction rate is about the same as their rate of conditioning, whereas in humans extinction occurs much more rapidly than acquisition. Indeed, as will be seen, under the condition of discontinuing the UCS, extinction occurs in most human subjects in one or two trials.

A number of studies have shown that extinction of the conditioned eyelid response is relatively slow in animals. Thus, Hilgard and Marquis (1936) found that of four monkeys who had their eyelids conditioned to a flash of light, three revealed no performance decrement whatever in 180 nonreinforced trials, while the fourth decreased to a level about 50% of the final acquisition strength. Subsequently, one of the monkeys was subjected to 720 extinction trials, at the end of which it was still responding at a 60% conditioned-response (CR) level (Hilgard & Marquis, 1936). A similar finding was obtained by Hilgard and Marquis (1935) in the case of eyelid conditioning in dogs, with extinction proceeding very slowly.

More recently experiments with rabbits (Gormezano, Schneiderman, Deaux, & Fuentes, 1962; Schneiderman, Fuentes, & Gormezano, 1962; Thomas & Wagner, 1964) have also revealed that extinction in these animals is a relatively prolonged process. For example, in a study in which extension of the nictitating membrane of the eye of the rabbit was conditioned to a tone, it was found that extinction took longer (560 trials) than was required to reach the acquisition asymptote (420 trials). Similarly, Bitterman and his colleagues have found that extinction of a classical aversive CR in the pigeon and the goldfish was slower than the original acquisition of the CR (Gonzalez, Longo, & Bitterman, 1961; Longo, Milstein, & Bitterman, 1962).

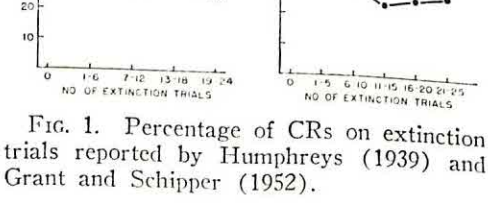

FIG. 1. Percentage of CRs on extinction trials reported by Humphreys (1939) and Grant and Schipper (1952).

In sharp contrast to these extinction findings obtained with animals are the data on extinction of the eyelid CR in humans. Figure 1 presents some typical data from two early investigations of the conditioned eye blink in humans (Grant & Schipper, 1952; Humphreys, 1939). In both of these studies the extinction procedure, as in the case of all of the animal studies, involved discontinuance of the UCS. The most significant thing to note about these curves is the rapidity with which extinction takes place. Less than 12 trials were required to reach a random level of responding in the Humphreys' (1939) study, while the curve of Grant and Schipper (1952) had reached this level by Trials 11–15. Not only do these graphs reveal extinction of the conditioned lid closure to be much faster in humans than animals, but also the ratio of extinction trials to acquisition trials is much smaller than in animals. Thus in these two studies the number of trials to extinction, 12–15, was only about 20–40% of the number of trials required to attain the acquisition asymptote, a finding quite typical of the data available in the literature. In animals, on the other hand, as many or even more trials are required to extinguish a response as were involved in the original conditioning.

But, as is now known from recent studies conducted in our laboratory, discontinuance of the air puff leads to even more rapid extinction in humans than was revealed by the data of these earlier experiments. The manner in which the curves in these studies were drawn actually masked the abruptness of the response decrement. The reasons for this are that (*a*) too large blocks of trials were employed in drawing the curves; and (*b*) the data for the first extinction trial was included in the first block of trials. Obviously, a response on the first extinction trial, occurring as it does before the first nonreinforcement, does not reflect the effects of this operation. The more appropriate procedure is to plot the data in terms of the number of previous nonreinforced trials. Moreover, graphing the curves in terms of single trials or blocks of two trials provides a more precise picture of the early course of the response decrement.

Figure 2 presents a recently obtained curve of extinction for human subjects drawn in this manner (Spence, Rutledge, & Talbott, 1963). Under this procedure the data for the first point (0) on the graph, which is the first trial on which no reinforcement is given, reflects performance after the last reinforced trial, while the data for Trials 1–2, 3–4, etc., represent the average frequency of response to the conditioned stimulus (CS) *after* the indicated number of nonreinforced trials. As may readily be seen, after only 2–4 extinction trials, the frequency of the CR has dropped to a random level.

This and other recent studies from our laboratory have clearly revealed that the rate with which the conditioned eye blink in humans decreases in strength under the extinction procedure of omitting the UCS is extremely rapid, not only as compared with animals, but also as compared with the rate of acquisition of the response in humans. Whereas from 50–80 trials are typically needed to condition the eyelid, extinction requires very few. Indeed, in the case of the above experiment more than half the subjects never gave a single CR following the first nonreinforced trial.

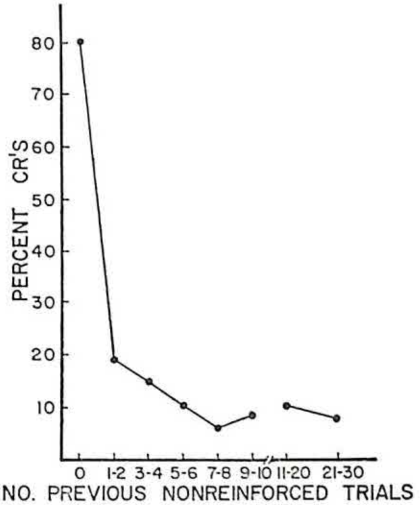

FIG. 2. Percentage of CRs during extinction as a function of number of previous nonreinforced trials in which UCS was omitted (Spence et al., 1963).

DRIVE LEVEL AND EXTINCTION OF EYELID CR

How is this striking difference in the extinction behavior of animals and man to be interpreted? One possibility, of course, is to assume that the quantitative properties of the effects of these two experimental operations, reinforcement and nonreinforcement, are very different in humans and simply assign appropriate parameter values to the rates of their development. But only the most naïve of theorists would proceed on such an assumption without considering the alternative possibility that the extreme rapidity of this extinction in humans reflects the operation of factors other than the development of a simple learning (nonreinforcement) effect.

Another possibility is that some kind of higher psychological process takes over control of the behavior during extinction in the case of the human subject. A number of early experiments (cf. Grant, 1939; Miller, 1939) had shown that instruction-produced sets can markedly facilitate or inhibit the level of response in such conditioning experiments. It seems entirely possible that self-instructions, of which the subject might or might not be aware, could play a decisive role in determining this extinction performance. The first idea that occurred was that the human subject readily recognizes the discontinuance of the noxious UCS, ceases to expect it, and thus reduces his level of emotional response (fear) and hence his drive level (D). The rapid extinction in humans could thus be accounted for in terms of a *sudden* drop in the level of D.

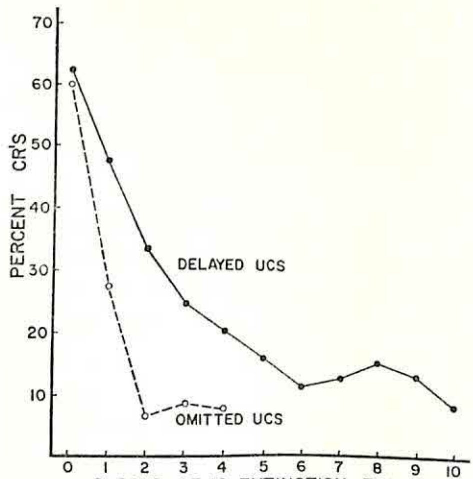

FIG. 3. Percentage of CRs on 10-trial extinction blocks for subjects extinguished with the UCS omitted and with the UCS delayed. (From McAllister, 1953.)

McAllister Experiment

An experiment that appeared to lend some suport to such a notion had been run some years earlier in our laboratory by McAllister (1953). After ascertaining that little or no eyelid conditioning occurs in human subjects when a CS–UCS interval of 2.5 seconds was employed, McAllister then demonstrated that CRs established at an optimal CS–UCS interval (.45 second) extinguished when the interval was extended to 2.5 seconds. Extinction of subjects under this latter method provides a test of the present drive hypothesis, for under this condition the UCS is continued, and thus the drive level of the subject is maintained throughout the extinction period. Accordingly, it would be expected that the performance curve would drop much more slowly under this condition than when the UCS is discontinued altogether.

As shown in Figure 3, McAllister's results confirmed this expected difference. The lower curve represents the performance of subjects extinguished without the UCS. The solid upper curve portrays the performance of subjects that had the delayed UCS. Both curves give the percentage of CRs that occurred in the interval 100–450 milliseconds following the onset of the CS for successive blocks of 10 trials. It is clearly evident that the decrement in performance developed much more slowly in those subjects that had the UCS present but delayed. In line with the above reasoning, McAllister interpreted this greater resistance to extinction as being due to the maintenance of the level of D when the delayed UCS was employed, whereas the level of D decreased when the UCS was discontinued. While this interpretation in terms of different levels of D nicely accounts for the performance difference between these two groups, it does not, it should be noted, explain why the extinction in human subjects is so precipitous, so rapid relative to that in animals which also had the UCS discontinued.

Reynolds Experiment

The inadequacy of this interpretation in terms of a difference in level of *D* only was further revealed by the findings of a subsequent study conducted for a doctoral dissertation by Reynolds (1958). While one part of Reynolds' study supported the findings of McAllister, another portion did not. McAllister's groups had, for another purpose, been conditioned with a partial reinforcement schedule (80%). Confirming McAllister, Reynolds found that when the original conditioning involved a partial reinforcement schedule (60%), the group with the delayed UCS again extinguished much more slowly than the group in which UCS was discontinued. This can be seen in Figure 4 in which the extinction curves for the partial groups (open circles) start at the 50% level.

In the case of subjects who acquired the CR under 100% reinforcement, however, Reynolds failed to find any difference whatever in extinction performance. As may be seen, the delayed CS–UCS interval group showed the same precipitous decrease in response strength as did the group that had no UCS in extinction. Note should be taken of the fact that the curves in this graph are of the earlier type that included the first nonreinforced trial in the first block of 10 trials. Omitting this trial, the level of response for the two groups on Trials 2–10 fell slightly below 20%. This finding with 100% reinforcement has since been replicated at two different levels of original conditioning in an experiment in which the ready signal was not employed (Spence et al., 1963). In both instances there was no difference in extinction rate between groups that had the UCS omitted and groups that had the UCS present but delayed during extinction.

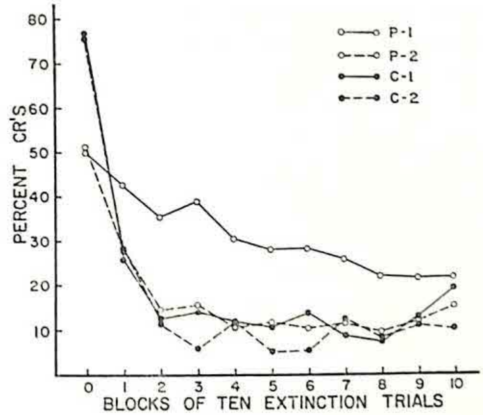

FIG. 4. Showing effects on extinction in standard situation of partial (P) and continuous (C) reinforced training trials and of delayed UCS (1) and omitted UCS (2) during extinction. (From Reynolds, 1958.)

DISCRIMINATION HYPOTHESIS: CONCEPT OF INHIBITORY SET

These findings clearly indicated that this simple drive interpretation was not adequate, and that it was necessary to appeal to some factor other than reduced drive level to account for both the facts of no difference in extinction rate and the extreme rapidity with which the CR extinguished. The clue followed in the attempt to ascertain what this factor might be was obtained by questioning subjects following extinction. In the Spence et al. (1963) study and in Goldstein's (1962) exploratory experiment it was found that practically all subjects reported that they recognized the change in conditions that occurred with the shift from acquisition to extinction. On the basis of this information it was hypothesized that human subjects, on noticing the change in stimulus conditions, become aware that their eyelid has been conditioned and thereupon more or less immediately adopt a set not to blink to the CS. Such an inhibitory set, it should be noted, would account for the sudden disappearance of the CR following consistent reinforcement.

Under partial reinforcement conditioning, on the other hand, the subject has experienced nonreinforcements on some portion of the acquisition trials (e.g., 50%). Hence the shift to extinction (i.e., 50%–0%) should not be as noticeable as the shift from 100% to no reinforcement. Presumably recognition of the shift to zero reinforcement in this circumstance would be delayed and possibly in some subjects would not occur at all. In this event the postulated inhibitory set would not be present, at least not during the early stages of extinction. Furthermore, with its absence the CRs would be able to occur and thus could reflect any differences in the drive level of the subjects that might be present. Accordingly, it would be expected that subjects in whom the UCS is discontinued would show a more rapid decrement than do subjects in whom the UCS is continued, but delayed. This predicted result is, it will be noted, in accord with the findings obtained with the partially reinforced groups of McAllister and Reynolds.

Finally, this hypothesis predicts the well-established finding that extinction of classical CR in humans is not nearly as rapid following partial reinforcement as after continuous. This interpretation of the difference in extinction performance following these different reinforcement schedules is, as will be recognized, a type of discrimination hypothesis. The present version, however, differs from previous ones in its introduction of a more specific hypothetical mechanism, the inhibitory set.

INTRODUCTION OF THE MASKING TASK

Our most recent investigations have been primarily directed at testing this theoretical interpretation of extinction by attempting to arrange experimental conditions that would reduce to zero, if possible, the discriminability of the change in conditions with shift from acquisition to extinction. To the extent that this objective could be achieved, the role of the hypothetical inhibitory set during extinction would be minimized. Presumably, if no discrimination occurred the set would not be operative, and extinction of the CR would be a function of simpler nonreinforcement effects (I_n), as in animals.

The effects of a number of changes have been investigated (Spence, Homzie, & Rutledge, 1964), the major one of which involved the use of a masking learning situation in terms of which the purpose of the experiment was specified. But first, the reasons for introducing this change: Questioning of the subjects following the completion of extinction in a recent study (Spence et al., 1963) revealed that 52% were able to recognize that the experiment was concerned with the conditioning of the eye-blink response. This percentage is to be compared with the findings of earlier studies that had involved conditioning, but not extinction. In such studies the percentage of subjects who were able to identify the experiments as involving conditioning had ranged around 5–10%. It is apparent from this that the added experience of extinction greatly increases the recognition by the subject of the true purpose of the experiment, thus enhancing the possibility that cognitive or mediating sets such as those postulated might play a critical role in determining performance during extinction.

A technique was sought, then, that would preclude the subject's recognizing that he was participating in a conditioning experiment even when extinction was attempted. It was also hoped that the procedure would decrease still further the likelihood that the changes in the stimulus events relative to extinction would be observed. The technique hit upon was that of embedding the conditioning experiment within the setting of another learning situation (Spence et al., 1964).

The learning task used for this purpose was the light-guessing, or probability-learning, task designed by Estes and Straughan (1954). This situation involved a centrally placed signal light which, when it came on, was the signal for the subject to anticipate within its duration (2 seconds) which of two small bulbs, one to the left, one to the right of the signal light, would subsequently light. The subjects were given a set of instructions to the effect that the experiment was concerned with the effects of distraction on performance in a difficult problem-solving situation. They were instructed that their task was to predict which of the two small lamps would light and to signify their prediction by pressing the push button located on the left or right arm of their chair. They were told further that distracting stimuli in the form of a tone and air puff to their eye would be given between their response of pressing a button and the lighting of one of the lamps. The onset of the tone coincided with the turning off of the signal light. The subjects were urged to predict to the best of their ability and to attempt to improve their prediction with practice. The small lamps lit according to a prearranged schedule. For half the subjects, the left lamp lit 7 times in a block of 10 trials, the right lamp 3 times. In the other half, these frequencies were reversed.

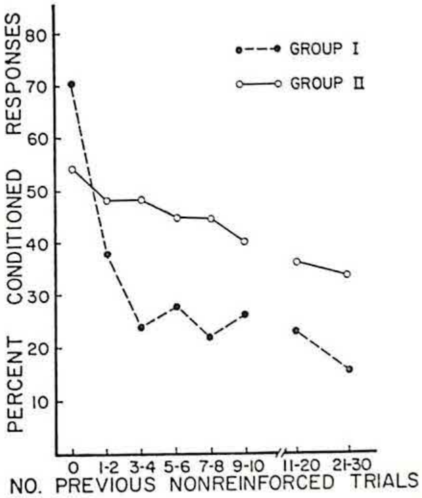

FIG. 5. Comparison of extinction rate in standard (Group I) and masking situations (Group II). (From Spence et al., 1964.)

EXPERIMENTS WITH THE MASKING TASK

The main finding of our first experiment employing this masking situation is shown in Figure 5. This graph compares the extinction of subjects (Group I) conditioned and extinguished with a delayed UCS in the standard conditioning situation with that of a group (II) in which the conditioning and extinction were embedded in the masking situation. As may be seen, the slopes of the two extinction curves differ markedly, and it would appear that the CR extinguishes much more slowly in the case of subjects run in the masking situation. Whatever its basis, the masking situation produced a most dramatic change in the behavior of human subjects. That it is related in some manner to the recognition on the part of the subjects of the change in stimulus conditions with shift to extinction is suggested by the results of a questionnaire given at the end of the experimental period. Thus 19 (76%) of the 25 subjects in Group I reported recognition of the shift in conditions, while only 1 (4%) of the subjects run in the masking situation was able to report the change.

A second study (Goldstein, 1962) was especially concerned with whether subjects, following 100% reinforcement in conditioning, would exhibit different extinction performances under the two procedures of omitting and delaying the UCS. It will be recalled that no difference was obtained between groups extinguished by these two methods under the standard conditions in the Reynolds (1958) study and Spence et al.'s (1963) replication of it. Moreover, it will be remembered that extinction was extremely rapid in both experiments. If, as assumed, the inhibitory set is not operative in the masking situation, a difference should now be obtained with these two extinction procedures involving as they do differences in the drive levels of the subjects.[2]

The results were in complete agreement with our assumption. As Figure 6 reveals, the group (C-1) that had the UCS present in extinction performed at a much higher level than the group (C-2) that had the UCS discontinued. The difference was highly significant ($p < .01$). The remaining two curves are for groups extinguished following partial reinforcement (60%). As in the McAllister (1953) and Reynolds (1958) studies, the group (P-1) that had the delayed UCS responded at a considerably higher level during extinction than Group P-2 in which the UCS was omitted.

The use of a more extended extinction period (100 trials) in the Goldstein (1962) study revealed, however, a new complication that had not been observed in previous data obtained with the delayed-UCS procedure. In the standard conditioning situation this method of extinction had always led to complete extinction of the response, that is, to the level of spontaneous blinking. As Figure 6 shows, however, the curves of groups extinguished with the delayed-UCS technique appear to be asymptotic at a much higher level (35–40%). A subsequent study (Spence & Rutledge, 1964) has shown that extinction curves obtained with this technique in the masking situation do not drop below this level (approximately 40%), even when 150 extinction trials were given. Examination of individual records indicated that this failure of the curves to drop to a random level of blinking was a consequence of the fact that approximately one-third of the subjects showed no response decrement, and some even showed a slight increase.

This finding suggests that the delayed-UCS technique is reinforcing for some subjects and hence is not a satisfactory procedure for maintaining drive level during extinction. The fact that it had not appeared in previous

---

[2] Attention should be called to the fact that a difference was obtained in the standard conditioning situation between these two extinction procedures in a study (Spence et al., 1964) in which the duration of the CS was not changed between the acquisition and extinction periods as had been the case in the earlier studies. The difference was ascribed to the lesser discriminability of the shift in conditions to extinction. Complicating these results, however, are the findings obtained by Price, Abbott, and Vandament (1965) of a difference under conditions similar to those of Reynolds (1958). It is possible that the relatively slow extinction of the delayed-UCS groups in the Price et al. study resulted from unidentified voluntary-form responders.

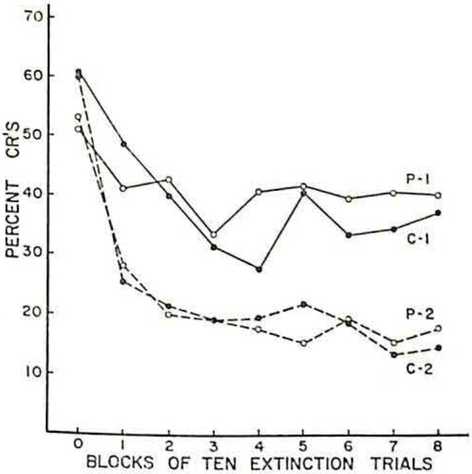

FIG. 6. Showing effect on extinction in masking situation of partial (P) and continuous (C) reinforced training trials and of delayed UCS (1) and omitted UCS (2) during extinction. (From Goldstein, 1962.)

studies using the standard situation can possibly be interpreted as resulting from the operation of the inhibitory set of the subject not to respond in this situation. With no inhibitory set present, as assumed, in the masking situation, the reinforcing effects of the delayed UCS would have an opportunity to be revealed.

But regardless of the explanation of this phenomena, it necessitated the finding of some other method than the delayed-UCS technique of maintaining the drive level of the subject during extinction. The procedure subsequently developed involved presenting the CS alone during extinction on a random half of the probability-learning (guessing-game) trials and the UCS alone on the other half. Thus the CS was never paired with the UCS, but the drive-maintaining puff was present during extinction. In order to maintain the average intertrial interval between successive presentations of the UCS in the original conditioning the same as during extinction, the paired CS and UCS were also presented on only a random half of the guessing-game trials during acquisition. On the remaining half of these trials neither the CS nor UCS occurred. This procedure thus gave a subject experience during acquisition with guessing-game trials in which no UCS was present and hence made the shift to extinction less discriminable.

The findings of an experiment (Spence, 1966) in which this new extinction technique was employed are shown in Figure 7. In addition to the subjects (Group I) extinguished in the manner just described two other groups were run. In Group II the CS was presented alone on half of the guessing trials with neither the CS nor the UCS being present on the other half. Since the UCS was not present in extinction, the drive level of Group II would, theoretically, be lower, and hence its performance should be lower than that of Group I. As the graph shows, this result was obtained ($p < .01$).

In the case of the third group the delayed-UCS technique was employed, thus providing an opportunity to com-

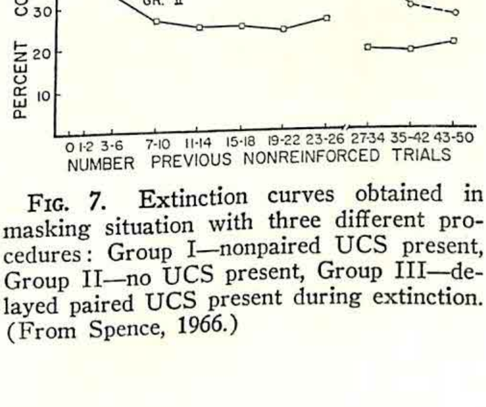

FIG. 7. Extinction curves obtained in masking situation with three different procedures: Group I—nonpaired UCS present, Group II—no UCS present, Group III—delayed paired UCS present during extinction. (From Spence, 1966.)

pare this previously used method of maintaining drive during extinction with the new procedure. In this group S_c and S_u were presented together at the long interval (2,500 milliseconds) on a random half of the guessing trials. In the other half of these trials neither stimulus was presented. Referring again to Figure 7, it is evident that the curve of Group III leveled off well above the random blink level of 18% response. On the other hand, the curve for the group extinguished with the new technique (Group I) dropped significantly below that of Group III ($p < .025$), reaching a level of about 27%. A still more recent study (Homzie, 1964) indicates that about 80 extinction trials are required to bring the curve obtained under this condition down to the spontaneous blinking rate. Since only 50 trials were needed in acquisition to attain asymptotic performance, it is apparent that, when cognitive factors are not operative and drive level is maintained, extinction of the eyelid CR in humans as in animals is somewhat slower than acquisition.

EXTINCTION AS A FUNCTION OF REINFORCEMENT SCHEDULES

The statement was made earlier that there were two major discrepancies between the extinction findings obtained with human and animal subjects in classical aversive conditioning. The second difference is that, whereas extinction in humans has heretofore consistently been found to be slower following a partial reinforcement schedule than after continuous reinforcement, the findings with animals reveal no such difference. For example, Thomas and Wagner (1964) have reported finding that extinction of the eye-blink response in the rabbit proceeded at the same rate following 50% and 100% reinforcement schedules in the original conditioning. Likewise, studies by Bitterman and his associates (Gonzales et al., 1961; Longo et al., 1962) have also failed to show a significant partial reinforcement effect in the extinction of classical defense responses in fish and pigeons. Finally it is of interest to note that Ross, Koski, and Yaeger (1964) have recently reported a study using human retardates (mean IQ, about 19) in which the partially reinforced subjects extinguished at the same rate as those that had received continuous reinforcement. It is apparent from these findings that the partial reinforcement effect in classical aversive conditioning is not present in organisms lacking effective higher psychological processes.

In terms of the present theoretical analysis the slower extinction following partial as compared to continuous reinforcement found heretofore in studies with humans is a result of the relatively slower and poorer discrimination of the shift to extinction on the part of the partial subjects. In the new masking situation this difference in discrimination should be minimal if not altogether absent; hence the partial reinforcement effect should not be evident. Examination of the two lowest curves in Goldstein's data (Figure 6) reveals that, in accord with this prediction, there is no difference in the curves of the partial and continuous groups (C-2, P-2) in which the UCS was omitted. In the case of the subjects that had the delayed UCS (C-1, P-1), the results are less clear-cut. In view of the possible reinforcing properties of this procedure, further research on this problem should employ the new extinction technique developed in connection with the masking task (Spence, 1966).

REFERENCES

AMSEL, A. The role of frustrative non-reward in non-continuous reward situations. *Psychological Bulletin,* 1958, **55,** 102–119.

ESTES, W. K., & STRAUGHAN, J. W. Analysis of a verbal conditioning situation in terms of statistical learning theory. *Journal of Experimental Psychology,* 1954, **47,** 225–234.

GOLDSTEIN, H. The effects of set factors and reinforcement schedules on the acquisition and extinction of the conditioned eyelid response. Unpublished doctoral dissertation, University of Iowa, 1962.

GONZALEZ, R. C., LONGO, N., & BITTERMAN, M. E. Classical conditioning in the fish: Exploratory studies of partial reinforcement. *Journal of Comparative and Physiological Psychology,* 1961, **54,** 452–456.

GORMEZANO, I., SCHNEIDERMAN, N., DEAUX, E., & FUENTES, I. Nictitating membrane: Classical conditioning and extinction in the albino rabbit. *Science,* 1962, **138,** 33–34.

GRANT, D. A. The influence of attitude on the conditioned eyelid response. *Journal of Experimental Psychology,* 1939, **25,** 333–346.

GRANT, D. A., & SCHIPPER, L. M. The acquisition and extinction of conditioned eyelid responses as a function of percentage of fixed ratio random reinforcement. *Journal of Experimental Psychology,* 1952, **43,** 313–320.

HILGARD, E. R., CAMPBELL, A. A., & SEARS, W. N. Conditioned discrimination: The effects of knowledge of stimulus relations. *American Journal of Psychology,* 1938, **51,** 498–506.

HILGARD, E. R., & MARQUIS, D. C. Acquisition, extinction and retention of conditioned lid responses to light in dogs. *Journal of Comparative and Physiological Psychology,* 1935, **19,** 29–58.

HILGARD, E. R., & MARQUIS, D. C. Conditioned eyelid responses in monkeys with a comparison of dog, monkey, and man. *Psychological Monographs,* 1936, **47** (2, Whole No. 212), 186–198.

HOMZIE, M. J. Separate phase differential eyelid conditioning within the context of a masking procedure. Unpublished doctoral dissertation, University of Iowa, 1964.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

HULL, C. L. *A behavior system.* New Haven: Yale University Press, 1952.

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reaction. *Journal of Experimental Psychology,* 1939, **25,** 141–158.

LONGO, N., MILSTEIN, S., & BITTERMAN, M. E. Classical conditioning in the pigeon: Exploratory studies of partial reinforcement. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 983–986.

MARQUIS, D. G., & HILGARD, E. R. Conditioned responses to light in monkeys after the removal of the occipital lobes. *Brain,* 1937, **60,** 1–12.

MCALLISTER, W. R. The effect on eyelid conditioning of shifting the CS-UCS interval. *Journal of Experimental Psychology,* 1953, **45,** 423–428.

MILLER, J. The effect of facilitatory and inhibitory attitudes on eyelid conditioning. Unpublished doctoral dissertation, Yale University, 1939.

PRICE, L. E., ABBOTT, D. W., & VANDAMENT, W. E. Effects of CS and UCS change on extinction of the conditioned eyelid response. *Journal of Experimental Psychology,* 1965, **4,** 437–439.

REYNOLDS, W. F. Acquisition and extinction of the conditioned eyelid response following partial and continuous reinforcement. *Journal of Experimental Psychology,* 1958, **55,** 335–341.

ROSS, L. E., KOSKI, C. H., & YAEGER, J. Classical eyelid conditioning of the severely retarded: Partial reinforcement effects. *Psychonomic Science,* 1964, **1,** 253–254.

SCHNEIDERMAN, N., FUENTES, I., & GORMEZANO, I. Acquisition and extinction of the classically conditioned eyelid response in the albino rat. *Science,* 1962, **136,** 650–652.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychological Review,* 1936, **43,** 427–429.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale University Press, 1956.

SPENCE, K. W. A theory of emotionally based drive and its relation to performance in simple learning situations. *American Psychologist,* 1958, **13,** 131–141.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

SPENCE, K. W. Extinction of the human eyelid CR as a function of presence or absence of the UCS during extinction. *Journal of Experimental Psychology,* 1966, in press.

SPENCE, K. W., HOMZIE, M. J., & RUTLEDGE, E. F. Extinction of the human eyelid CR as a function of the discriminability of the change from acquisition to extinction. *Journal of Experimental Psychology,* 1964, **67**, 545–552.

SPENCE, K. W., & RUTLEDGE, E. F. The effect of overlearning on rate of extinction of the eyelid CR. *Psychonomic Science,* 1964, **1**, 315–316.

SPENCE, K. W., RUTLEDGE, E. F., & TALBOTT, J. H. Effect of number of acquisition trials and the presence or absence of the UCS on extinction of the eyelid CR. *Journal of Experimental Psychology,* 1963, **66**, 286–291.

THOMAS, E., & WAGNER, A. R. Partial reinforcement of the classically conditioned eyelid response in the rabbit. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 157–158.

(Received July 12, 1965)

Psychological Review
1966, Vol. 73, No. 5, 459-477

PARTIAL REINFORCEMENT: A HYPOTHESIS OF SEQUENTIAL EFFECTS[1]

E. J. CAPALDI

University of Texas

A hypothesis which attempts to account for extinction following partial reinforcement in terms of certain sequential aspects of nonreinforced and reinforced trials was considered. The central assumptions of the hypothesis are that nonreinforced trials occasion a specific and distinctive internal stimulus and that this stimulus is progressively modified by successive nonreinforcements. Stimulus modification was indicated as exerting its effects on extinction through 3 sequential variables: N-length, number of different N-lengths, and number of occurrences of each N-length. The operation of these variables and the stimulus modification mechanism was elaborated by considering certain salient and representative partial reinforcement investigations. This detailed survey of critical experimental findings suggested that the better known nonsequential variables which form the basis of most current partial reinforcement hypotheses exercise little or no control over extinction and that a theory of extinction must be based on a sequential analysis. Experimental evidence was considered which suggests that the modification mechanism may be applicable to all partial reinforcement situations, whether classical or instrumental, appetitive or aversive. The suggestion was also advanced that the present hypothesis may be able to deliver certain types of rate-of-extinction predictions. These particular rate predictions, while relatively primitive, do appear nevertheless to be something of an improvement over the current greater-than type prediction.

Reinforcement (R) and nonreinforcement (N) or, more precisely, the organism's reactions to these, appear to occasion potent internal stimuli which regulate behavior in a variety of learning situations. The exclusive focus here, however, will be upon the role of these stimuli in governing partial reinforcement (PR) effects. Central to the present treatment is the assumption that the isolation of lawful relationships in PR is dependent upon a detailed and exhaustive examination of certain aspects of the sequence of N and R trials, a view which current major PR hypotheses reject, either explicitly or implicitly. The present emphasis on sequential effects is reflected only indirectly, and indeed sometimes not at all, by terms currently employed to describe PR schedules, such as percentage, pattern, or fixed ratio of reinforcement, etc., these terms having originated in different and sometimes contradictory theoretical frameworks. While such terms are not regarded as being particularly useful, either for present purposes or generally, they will be employed in what follows solely to provide a terminological bridge to the more appropriate sequential variables.

PR VARIABLES

The difficulty of achieving adequate experimental control in PR situations approaches the excessive; indeed, most PR investigations admit of any number of interpretations if only because the groups compared differ along so many dimensions. The general failure

---

[1] This work was supported in part by National Institute of Child Health and Human Development Research Grant HD 00949-01, -02.

TABLE 1

TWO SCHEDULES OF REINFORCEMENT ATTEMPTING TO ILLUSTRATE THE MAJOR SEQUENTIAL AND NONSEQUENTIAL VARIABLES IN THE PR SITUATION

Schedule	Trial					
	1	2	3	4	5	6
A	N	N	N	N	R	R
B	R	N	R	N	N	R

to come to grips with this problem has acted as an impediment to adequate and effective theorizing in the PR area. At the same time, it has probably led to the proliferation of PR hypotheses since, in many instances, the introduction of a "new" hypothesis amounts to little more than an arbitrary emphasis upon this or that PR variable to the exclusion of others. A more general orientation seems appropriate and for that reason the present hypothesis is considered in relation to the major variables in the PR area.

The two schedules of reinforcement appearing in Table 1 are typical of those employed in a wide variety of PR investigations. Those variables independent of the sequence of N and R trials are here termed nonsequential variables, those dependent upon this sequence, sequential variables. Three of the four major nonsequential variables are illustrated in the above schedules: percentage of reinforcement (33% versus 50%), number of N trials (4 versus 3), and number of R trials (2 versus 3). The fourth, number of training trials, is occasioned when an investigator attempts, as infrequently occurs, to vary percentage while holding constant the number of N or R trials. The literature contains many investigations in which some single aspect of the distribution of N and R trials was varied with percentage of reinforcement and total number of acquisition trials held constant. Such investigations have usually varied number of transitions from N trials to R trials (NR transitions) and have been termed patterning investigations (e.g., Grosslight, Hall, & Murnin, 1953).[2]

It is clear that in a patterning investigation, as defined above, the nonsequential variables do not shift in level. While this may seem something of an experimental advantage, and for certain gross purposes it is, the overall gain is slight since the sequential variables are quite numerous and usually all of these have shifted in level along with the one selected for experimental attention. There are illustrated in Table 1, for example, four varieties of trial-to-trial transition alone: NR (1 versus 2), NN (3 versus 1), RR (1 versus 0), and RN (O versus 2). Other sequential variables illustrated include the probability of an R trial following an N trial (.24 versus .66) and the remaining similar probabilities, N following N, R following R, and N following R.

According to the present hypothesis, the major determiners of resistance to extinction (Rn) are the three length variables: N-length, number of different N-lengths, and the number of occurrences of each N-length. The number of N trials occurring in succession without interruption by an R trial defines N-length. Schedule A in Table 1 contains a single N-length of 4, occurring but once; Schedule B contains two different N-lengths of 1 and 2, respectively, each occurring only once.

---

[2] Investigations concerned with NR transitions were inspired in the main by an historically earlier form of the present hypothesis, the Hull (1952) and Sheffield (1949) aftereffects hypothesis. Many of the assumptions of the Hull-Sheffield hypothesis have been discarded, modified, or extended (Capaldi, Hart, & Stanley, 1963; Capaldi & Spivey, 1963, 1964b; Capaldi & Stanley, 1963).

THEORETICAL BACKGROUND

Stimulus Modification

Experimental evidence indicates that reinforcement and nonreinforcement occasion distinctive stimuli, represented respectively as S^R and S^N (e.g., Bloom & Capaldi, 1961; Capaldi & Senko, 1962; Capaldi & Stanley, 1963). The basic theoretical assumption here is that S^N is modified by successive N trials and thus is the mechanism underlying the N-length variables. In more specific terms, the stimulus consequent of a single N trial, or S^{N_1}, differs either quantitatively, qualitatively, or both from the stimulus consequent of two successive N trials or S^{N_2}. The modification process occurs according to a simple growth function so that $\Delta S^N \rightarrow 0$. Stimulus modification continues either until an R trial occurs, in which event S^N is conditioned to the instrumental response and is replaced by S^R, or until the limit of the modification process is reached, a possible event given great numbers of consecutive N trials as in extinction.

The primary purpose of this paper is to demonstrate that the major findings in the area of percentage and pattern of reinforcement can be explained on the basis of stimulus modification. While we shall soon turn to the general literature, it is best to avoid doing so at present. In all but a few such investigations, variation in N-length, upon which stimulus modification depends, has simply been the consequence of the manipulation of some other PR variable. The better procedure seems to be to cover first, in rather extensive detail, two investigations which were specifically designed with N-length in mind (Capaldi, 1964). When originally published, the full theoretical implications of these investigations could not be spelled out for reasons of space limitation. Here, however, detailed coverage of these investigations seems almost mandatory, first, to make evident certain general principles which should transfer without difficulty to the remaining PR literature, and second, to provide a concrete illustration of the way the hypothesis is to be applied in specific circumstances, which should facilitate the subsequent inspection of the general literature as well as indicating how the hypothesis is to be tested.

Single Values of N-length

One of the simplest theoretical cases involves presenting the same number of reinforcements for only a single value of S^N. Two groups, one trained under the schedule NR, the other under the schedule NNR, illustrate the case; one reinforcement is provided, respectively, for S^{N_1} and S^{N_2}. As suggested previously, however, the implications of a fixed-ratio investigation of this kind are far too indeterminate to be particularly useful for theoretical purposes; the groups differ not only in N-length but in terms of number of N trials, training trials, percentage of reinforcement, NN transitions, etc.

A theoretically more useful experimental situation is provided by employing intertrial reinforcement (ITR). The ITR procedure consists of placing the subject (*S*) in the baited goal box during the intertrial interval. As indicated elsewhere (Capaldi, Hart, & Stanley, 1963; Capaldi & Spivey, 1963, 1964), an ordinary R trial is assumed to serve two general functions; (*a*) It conditions whatever stimuli are operative including S^N to the instrumental response (reinforcement function), and (*b*) it replaces the currently operative S^N (i.e., S^{N_5}, etc.) with S^R (replacement function). Since ITR occurs some time after the locomotor response was executed, it can be assumed not to serve the reinforcement function.

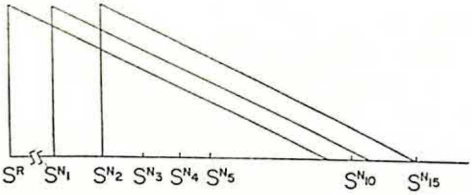

FIG. 1. Assumed theoretical consequences of delivering the same number of reinforcements at three different values of the aftereffect, S^R, S^{N_1}, and S^{N_2}. (The solid vertical lines above S^R, S^{N_1}, and S^{N_2} represent the habit strength available at each of these stimulus points. The generalization of habit is also shown. The modification undergone by the nonreinforced aftereffect is shown along the abscissa.)

However, like an ordinary R trial, it can serve the replacement function. Experimental evidence has confirmed these assumptions in connection with ITR (Capaldi et al., 1963; Capaldi & Spivey, 1963, 1964a).

Consider three groups, all of which are trained under the five-trial schedule, R, N_1, N_2, N_3, R. Let ITR occur following N_3 in Group S^R, following N_2 in Group S^{N_1}, and following N_1 in Group S^{N_2}. According to the assumptions considered above, in Group S^R ITR will replace S^{N_3} with S^R; however, S^{N_3} will not be conditioned to the locomotor response whereas S^R will be so conditioned on the final R trial. Similarly, in Group S^{N_1}, S^{N_1} will be conditioned to the locomotor response, and in Group S^{N_2}, S^{N_2} will be so conditioned. Note that by employing ITR it has proved possible to vary N-length while holding all nonsequential variables constant, and, provided one considers the schedule of reinforcement per se, all sequential variables are held constant as well. The assumed theoretical consequences of the above manipulation are shown in Figure 1. For illustrative purposes only, a rather large growth factor ($F = 1/10$) is used in the positive growth function employed to estimate stimulus modification. Stimulus modification is shown along the abscissa of Figure 1. The solid vertical lines in Figure 1 represent the amount of habit strength for the instrumental response possessed by S^R, S^{N_1}, and S^{N_2}, respectively. The lines are of equal height because each group received the same number of reinforcements for the particular stimulus value conditioned. The important feature deserving attention in Figure 1 is that the higher the value of S^N gaining habit strength for the instrumental response, the greater the amount of generalized habit strength provided to still higher values of S^N (i.e., $S^{N_{10}}$, etc.). Since, in extinction, the modification process continues without reversal, it follows that Rn will increase with an increase in the value of S^N conditioned to the instrumental response. An ITR investigation similar in its logical features to the one described above gave the theoretically expected result (Capaldi, 1964). Thus, it appears that *Rn increases as N-length increases* or, in theoretical terms, that Rn increases with an increase in the value of S^N conditioned to the locomotor response.

Multiple Values of N-length

Two other variables were mentioned as determiners of Rn: the number of times a given N-length occurs and the number of different N-lengths. On the assumption of a gradually increasing habit strength, it is clear that Rn will increase with the number of times a given S^N is reinforced. Moreover, when a schedule includes several different N-lengths, as do what are commonly termed irregular schedules or variable ratio schedules, several different values of S^N will be conditioned to the response, and it could be expected on theoretical grounds that the respective generalization tendencies will summate (Hull, 1943), thereby further increasing Rn. The investigation described in this section is concerned with these predictions.

In a recent investigation (Capaldi, 1964), the summation assumption was tested employing a single N-length group (G3) and a multiple N-length group (G123), both trained in a runway each day under the six-trial 50% schedules appearing in Table 2. Since either 24, 60, or 120 acquisition trials were given prior to extincton, there were a total of six groups. Note that since percentage of reinforcement and total number of acquisition trials were equated for the single and multiple N-length groups, the investigation would be referred to, employing typical terminology, as a patterning investigation. That it contains many features not incorporated into the typical patterning investigation will become apparent as the discussion develops, especially in later sections of the paper.

A convincing test of the hypothesis that generalized S^N tendencies summate to increase Rn requires that two conditions be met. First, the highest reinforced value of S^N must be the same for all groups (e.g., S^{N_3} in the present case). A longer N-length in the single N-length group might result in greater Rn in that group relative to a multiple N-length group simply because N-length is longer, and the experimental results would be indeterminate. A longer N-length in the multiple N-length group, on the other hand, would certainly result in greater Rn for that group relative to the single N-length group for reasons of N-length *and* summation, and the experimental results would again be indeterminate.

The second condition is that the highest value of S^N (e.g., S^{N_3} in the present case) must possess equal habit strength in both the single and multiple N-length groups. If such were not the case, the stronger S^{N_3} in G3 might provide more generalized habit strength to higher valued S^Ns than the combination of the three weaker S^Ns in G123 (S^{N_1}, S^{N_2}, S^{N_3}). Indeed, as will be seen, this prediction was tested at the 24-trial level. Notice that in a 4-day block (24 acquisition trials), S^{N_3} occurs *four* times in G3 and only *twice* in G123. Thus we are still faced with the original issue which is, is it possible to equate habit strength at S^{N_3} yet avoid introducing a host of other variables—training level, etc.—which would tend to impose their own interpretations limitations on the results?

The solution to this difficulty involves recognizing that habit strength mirrors number of reinforcements only to a point. That is, once habit becomes asymptotic, further reinforcements will not further increase habit

TABLE 2

REWARD SCHEDULES EMPLOYED IN THE CAPALDI (1964) INVESTIGATION

	G3						G123						
	Trial							Trial					
Day	1	2	3	4	5	6	Day	1	2	3	4	5	6
1	R	N	N	N	R	R	1	R	N	N	N	R	R
2	R	R	N	N	N	R	2	R	R	N	N	N	R
3	R	N	N	N	R	R	3	N	N	R	N	R	R
4	R	R	N	N	N	R	4	N	R	N	N	R	R

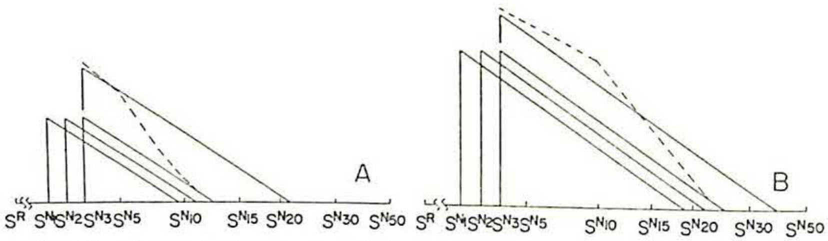
A
B
S^R S^{N_1} S^{N_2} S^{N_3} S^{N_5} $S^{N_{10}}$ $S^{N_{15}}$ $S^{N_{20}}$ $S^{N_{30}}$ $S^{N_{50}}$
S^R S^{N_1} S^{N_2} S^{N_3} S^{N_5} $S^{N_{10}}$ $S^{N_{15}}$ $S^{N_{20}}$ $S^{N_{30}}$ $S^{N_{50}}$

FIG. 2. Theoretical curves. (The modification undergone by the nonreinforced aftereffect as a function of successive nonreinforcements is shown along the abscissa. Only salient stimulus points, e.g., S^{N_1}, $S^{N_{15}}$, etc., are shown. Figure A depicts the theoretical situation following 24 trials of acquisition; Figure B, following 60 trials. The solid vertical lines represent the amount of habit strength available at each value of S^N conditioned during acquisition training. The lower vertical line represents the amount of habit strength available at S^{N_3} for G123; the lower line plus the line above represents the amount of habit strength available at S^{N_3} for G3. The generalization of the various habit strengths is also shown. The dotted line shows the effects of habit summation in G123. The width of the generalization was obtained in each instance by multiplying the height of the vertical line for habit by 1.5.)

strength. In line with this consideration, in G3 S^{N_3} received either 4 reinforcements (24 trials), 10 (60 trials), or 20 (120 trials) prior to extinction, while in G123 the corresponding values for S^{N_1}, S^{N_2}, and S^{N_3} were 2 each at the 24–trial level, 5 each at the 60–trial level, and 10 each at the 120–trial level. The growth factor (F) chosen for habit was .25 and for stimulus modification .05, these F values being determined, as will be seen, on the basis of previous investigations. With $F = .25$, habit becomes nearly asymptotic following 10 reinforcements. Thus, at the 120–trial level, habit strength at S^{N_3} will be equal or nearly so for G3 and G123, that is, asymptotic, and the summation assumption can be tested. The groups receiving 24 or 60 acquisition trials are included because they place additional constraints upon the hypothesis. Most importantly, if the assumptions of an F or .25 and .05 for habit and stimulus modification, respectively, are in gross error, the hypothesis very likely will not fit simultaneously the data obtained from the three different training levels.

Theoretical curves illustrating the situation following 24 acquisition trials are presented in Figure 2A. Since G123 received two reinforcements each at S^{N_1}, S^{N_2}, and S^{N_3}, each has 43.75 units of habit strength, while in G3, S^{N_3} has 68.36 units (four reinforcements). The dotted line in Figure 2A shows the effects of summation in G123; it can readily be seen that number of reinforcements at S^{N_3} should outweigh summation as a determiner of Rn at the 24–trial level. Summation was estimated employing the equation suggested by Hull (1943). According to the theoretical curves, the two groups should run about equally rapidly in the very early extinction trials, with G123 showing more decrement than G3 later in extinction. That is, the extinction curves should begin to diverge following the early extinction trials. The experimental results confirmed these predictions.

Figure 2B illustrates the theoretical situation following 60 acquisition trials and suggests that G123 should, if anything, be faster than G3 in the early extinction trials but slower later in extinction. Thus, the extinction curves should converge early in extinction followed relatively late in extinction by divergence. While convergence was obtained, late-trial divergence was not, presumably because sufficient extinction training was not given (30 trials). Further, as Figure 2B suggests, differences between the groups were small and nonsignificant.

Theoretical curves are not needed at the 120-trial level since matters are clear without them. At 120 trials habit strength at S^{N_3} will be approximately equal for G3 and G123 and, moreover, G123 will in addition have near asymptotic habit tendencies available from S^{N_1} and S^{N_2}. Thus, while the groups are not expected to differ in the very early extinction trials, as extinction progresses G123 should show less decrement over trials than G3, that is, the extinction curves should diverge over trials. These predictions were confirmed by the experimental results which demonstrate, as predicted by the hypothesis, that the habit tendencies available from several S^Ns summate to increase Rn.

The investigations considered above suggest that N-length, number of different N-lengths, and number of occurrences of each N-length are critical variables determining Rn. Keeping them in mind, it is now possible to turn to the major findings on percentage and pattern of reinforcement. Doing so will serve two important purposes. Most obviously, it can be determined if the N-length variables can be employed to explain these findings. It can also be determined if the results of the investigations just considered are in any way incompatible with those contained in the general literature.

MAJOR PR FINDINGS

Small Numbers of Acquisition Trials: Pattern and Percentage

With a single exception (Spivey, 1964), none of the experiments to be considered was specifically designed to test the present hypothesis although some were designed to test a more primitive form of it. It can be expected, therefore, that certain interpretational difficulties will be encountered, the most common being that one group usually received three or four times the number of reinforcements at some lower value of S^N than another received at some higher value. Since both factors, number of reinforcements and S^N value, operate to increase Rn, it will prove necessary in the small-trials case to present a somewhat detailed analysis employing theoretical curves. Inspection of those curves may be facilitated by bearing in mind a general rule-of-thumb from the hypothesis for the small-trials case: Number of reinforcements will usually outweigh value of S^N as a determiner of Rn.

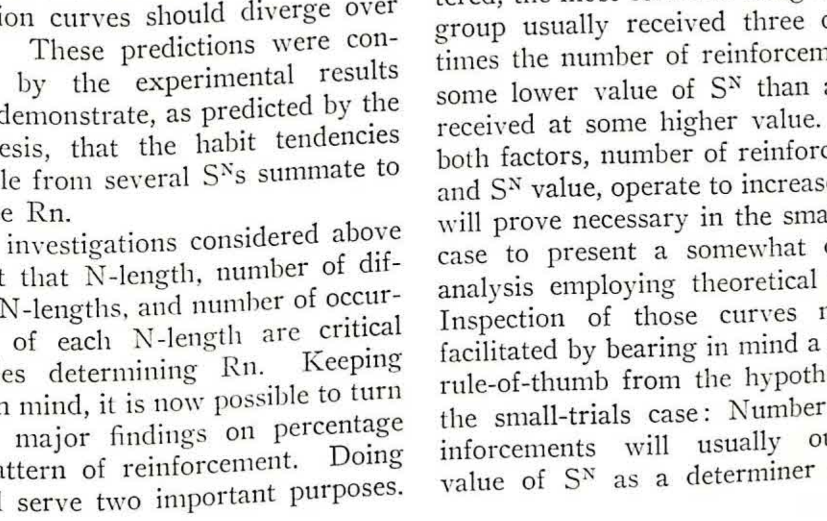

FIG. 3. The curves on the left show the assumed theoretical consequences of delivering 8 reinforcements at S^{N_1} versus 2 reinforcements at S^{N_2}, whereas the curves on the right show the effects of 12 reinforcements at S^{N_1} versus 3 at S^{N_2}.

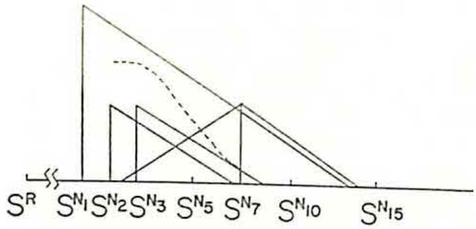

Fig. 4. Curves illustrating the theoretical situation in Bacon's (1962) investigation. (The 70% group received three reinforcements at S^{N_1}, the 50% group, one at S^{N_2} and one at S^{N_3}, whereas the 70% group received one reinforcement at S^{N_7}.)

That is, if the desire is to produce relatively great Rn employing a relatively small number of acquisition trials, reinforcements should be concentrated at a single value of S^N (e.g., see G3 versus G123 at 24 trials in the previous section). The theoretical curves to be presented employ the same set of specific assumptions as previously, that is, $F = .25$ for habit and .05 for stimulus modification. The applicability of these specific assumptions to experiments from a variety of laboratories seems encouraging. All small-trial investigations to be considered employed the runway.

Pattern of reinforcement. In two separate small-trial patterning investigations (18 and 27 trials, respectively), it was found (Capaldi & Hart, 1962) that a 50% alternating (A) schedule of N and R trials produced greater Rn than a 50% irregular (I) schedule. In the 18-trial investigation, the A group received eight reinforcements at S^{N_1}, the I group received two at S^{N_2}. In the 27-trial investigation, the A group received twelve reinforcements at S^{N_1}, the I group received three at S^{N_2}. Theoretical curves from which the greater Rn of the A group, relative to that of the I group, may be deduced are shown in Figure 3.

Percentage of reinforcement. Following 10 acquisition trials, Bacon (1962) found Rn to be an increasing function of percentage of reinforcement. An examination of Bacon's reward schedules (personal communication) indicated that his 70% group received three reinforcements at S^{N_1}, his 50% group received one at S^{N_2} and one at S^{N_3}, and his 30% group received one at S^{N_7}. Theoretical curves for these groups are shown in Figure 4. For the 30% group, the generalization of habit tendencies is shown extending in both directions along the continuum, and for the 50% group the summation of the two habit tendencies is shown (dotted line). In examining Figure 4, it should not be inferred that the present hypothesis expects either that the 30% group would not run in the early extinction trials or that the 50% group would not run in the later ones. Running would be maintained to some extent on these trials by sources of habit strength not particularly germane to the present theoretical analysis (e.g., alley stimuli). Figure 4 shows clearly,

TABLE 3

Reward Schedules Employed in the Spivey (1964) Investigation

Schedule	Trial										S^N Reinforced
	1	2	3	4	5	6	7	8	9	10	
30-1	R	R	N	N	N	N	N	N	N	R	1 at S^{N_7}
30-3	N	N	R	N	N	R	N	N	N	R	2 at S^{N_2}, 1 at S^{N_3}
70-1	R	R	R	N	N	N	R	R	R	R	1 at S^{N_3}
70-3	R	N	R	R	R	N	R	N	R	R	3 at S^{N_1}

however, that the 70% group should show greater Rn than either the 50% or 30% groups. Presumably, the superiority of the 50% group in the early extinction trials outweighed the superiority of the 30% group in the later extinction trials to produce overall greater Rn. However, Bacon did not present his results in a form which allows a determination on this score. That some such complicated Trials × Percentages interaction may have in fact occurred in Bacon's study is suggested not only by the theoretical curves in Figure 4 but also by the results of a recent doctoral investigation by Spivey (1964).

Spivey employed the same 70% schedule (70–3) and the same 30% schedule (30–1) employed by Bacon, employing two other schedules as well (70–1 and 30–3). The four schedules are shown below in Table 3.

The schedules are designated on the basis of percentage reinforcement (70 versus 30) and number of reinforcements (1 versus 3) for some value of S^N. Spivey found Group 70–3 to show greater Rn than Group 30–1, these results replicating Bacon's. On the other hand, Group 30–3 showed significantly greater Rn than Group 70–1. The upshot was that, combining groups, percentage of reinforcement had no effect on Rn while number of different S^Ns reinforced had a significant effect. In addition, there was a significant interaction between trials and number of reinforcements for S^N. For example, Group 70–3, while showing a greater overall Rn than Group 30–1, extinguished at a faster rate over the first 10 trials of extinction, that is, the superiority of Group 70–3 relative to Group 30–1 occurred mainly in the early extinction trials, a result predicted in Figure 4. Indeed, by the tenth extinction trial Group 30–1 was actually running somewhat faster than Group 70–3. It should be clear, without presenting further theoretical curves, that the results of the Spivey investigation are consistent with the present hypothesis, as are those of the Bacon investigation.

Extended Acquisition Training: Pattern and Percentage

As previously indicated, following considerable acquisition training it may reasonably be assumed in most cases that all or nearly all values of S^N conditioned to the response have an asymptotic, or a nearly asymptotic, capacity to evoke the locomotor response. Thus, in the extended-trial case, as in any equal habit strength case, it is most easy to formulate a rule of thumb for determining which of several groups should show the greater Rn. Actually, it is the reverse of the rule applied to the small-trial, or unequal habit strength, case. For small trials, it was indicated that the best way to increase Rn was to concentrate reinforcement at a single value of S^N. For extended training, the hypothesis recommends the opposite procedure: Spread reinforcements over a number of different values of S^N, particularly high values of S^N. The two rules of thumb taken together suggest that the schedule which produces the least Rn for small numbers of training trials should produce the greatest Rn for extended numbers of training trials, and vice versa. As will be shown below, this general observation appears to be consistent with all available evidence.

Pattern of reward. An example of the Schedule × Number of Acquisition Trials interaction has already been covered in connection with pattern of reward. It may be recalled from a previous section that G3 showed greater Rn than G123 at 24 trials and less Rn at 120 trials. Moreover, following 120 trials Tyler, Wortz, and Bitterman which is central to the present treatment. Moreover, that nonreinforcement is "registered" independently of a tendency toward extinction is one important aspect of the stimulus modification assumption.

COMMENTS ON MAJOR FINDINGS

Every effort has been made to cover a representative sample of PR findings, with respect both to training level and schedule of reinforcement. Moreover, practically every investigation included in the literature survey has at one time or another been replicated. In the few cases where this was not so, the unreplicated findings are clearly consistent with the remaining literature, if not actually implied by that literature. What have been called major findings are so, then, in two senses, representativeness and replicability.

While the N-length variables appear to have survived the test of the major findings, it seems unlikely on the basis of these findings that the nonsequential variables exert any but the mildest of influence on Rn. For example, increased percentage of reinforcement or increased number of N trials sometimes accompanied increased Rn (Spivey, 1964; Weinstock, 1954, 1958) and sometimes accompanied decreased Rn (Bacon, 1962; Spivey, 1964). Moreover, Rn has been shown to vary when both percentage of reinforcement and number of N trials remained constant (Capaldi, 1958; Gonzalez & Bitterman, 1964; Tyler et al., 1953). Interestingly, even a restricted claim for a nonsequential variable encounters overwhelming difficulties when tested against the experimental evidence. For example, it has been suggested (e.g., Bacon, 1962; Grant & Schipper, 1952) that at least for small numbers of training trials Rn increases as number of R trials increase. This restricted generalization is inconsistent with the results of small trial patterning investigations (e.g., Capaldi & Hart, 1962) and, among percentage investigations, with Spivey (1964) where following 10 acquistion trials Rn was greater for a group given three R trials than for a group given seven R trials (30–3 versus 70–1). Thus, it would appear that the conclusions available from the two investigations covered in detail previously (Capaldi, 1964) in which N-length, number of different N-lengths, and number of occurrences of each N-length were varied are adequately supported by the results of investigations contained in the general literature which, for the most part, were neither concerned with the N-length variables nor designed with these variables in mind. Moreover, it would appear on the basis of available evidence that it is possible to systematically rule out of consideration as major determiners of Rn practically all of the numerous variables contained in the typical schedule of partial reinforcement except the N-length variables.

The exceptions, if such actually exist, are number of RR and number of RN transitions. According to a more expanded form of the present hypothesis, these variables *should* exert some influence on Rn. Our impression is that an especially detailed analysis would show both variables to exert some influence on Rn under PR schedules. The need for "an especially detailed analysis" is occasioned by the fact that the far more powerful N-length variables appear to have obscured the effects of these "milder" transition variables in practically every PR investigation so far reported. Be that as it may, to attempt to treat RR or RN transitions at the present time would occasion a highly speculative discussion, one which for a variety of reasons is perhaps best avoided.

FURTHER IMPLICATIONS

Experimental Situations

Ordinarily, PR hypotheses tend to deal with a relatively restricted class of experimental situation, for example, the present one has tended to deal mainly with appetitive conditioning in the runway. Such selectivity sometimes occurs because the mechanisms of a given hypothesis plainly do not extend beyond the experimental situation chosen and sometimes for other reasons. The most obvious and important issue resulting from this selectivity concerns the number of different sets of laws which need to be formulated to deal with all PR situations (cf., for example, Spence, 1960). Hopefully, one set of laws will prove sufficient.

The stimulus mechanism of the present hypothesis seems to have a broad application, one which, at least in principle, could apply to all PR situations. Since the hypothesis is proposed as being as applicable, say, to classical aversive conditioning as to the instrumental runway situation, the primary issue to be faced at present is how well the hypothesis fits the data in experimental situations other than the runway. Our impression is that the fit is good.

For example, Jensen (1964) has shown that N-length and number of different N-lengths affect Rn in the free operant situation (rat subjects (*S*s) and bar pressing). It should be noted that Jensen's free operant results are consistent with those of Gonzalez and Bitterman (1964) who employed discrete bar presses, although the Jensen investigation was by far the more elaborate. Jensen employed five levels of N-length (1, 15, 20, 50, and 100) and three levels of occurrence of each N-length (1, 3, or 7) and found both variables to have a significant effect on Rn.

Unfortunately, however, the available data from many experimental situations are far too unsystematic to allow as rigorous a test of the present hypothesis as would seem both desirable and necessary. For example, Wagner, Siegel, Thomas, and Ellison (1964) have recently indicated that there are no systematic data available on the extinction of classical rewarded responses in infrahuman *S*s as a function of percentage of food reinforced trials during acquisition. They conducted such an experiment with dogs, comparing 50% with 100% reinforcement, and suggested that the extinction findings, while not compatible with alternative PR hypotheses, could be understood on the basis of greater generalization decrement in the 100% group ". . . as traces of nonreinforcement rather than reinforcement are first introduced for this group into the stimulus complex [p. 358]." Consider, too, a finding reported by Grant, Riopelle, and Hake (1950) employing human *S*s in an aversive eyelid conditioning situation. Grant et al. found alternating schedules produced greater Rn than irregular schedules, a finding, which it will be remembered, similar to that reported by Capaldi and Hart (1962) following a small number of acquisition trials in the runway. The Grant et al. finding could offer strong support for the present hypothesis; indeed, we are aware of no other hypothesis from which it can be deduced. It could offer such support but, in one important respect, it does not. This is so because there is no direct way of determining if the number of acquisition trials employed by Grant et al. was small for the eyelid situation since, unfortunately, a comparable investigation employing more extended acquisition training does not seem to be presently available. If an extended trial investigation showed irregular patterns produced greater Rn than alternating patterns, the probability would be increased that the present hypothesis is applicable to the eyelid situation and perhaps to other aversive situations as well.

Studies from the Bryn Mawr laboratory employing the fish and the bird have shown that following extended acquisition training (400 or more trials) long N-length produces greater Rn than short N-length (Gonzalez, Eskin, & Bitterman, 1963; Gonzalez, Graf, & Bitterman, 1965), but following a brief acquisition period (50 trials) long and short N-length produce about the same Rn (Gonzalez, Behrend, & Bitterman, 1965; Roberts, Bullock, & Bitterman, 1963). These results for fish and bird are similar to those for the rat considered in detail earlier in showing that the greater Rn associated with longer N-lengths occurs only after considerable acquisition training, that is, in all three organisms, fish, bird, and rat, there appears to be a similar schedule of Reward × Acquisition Trial interaction. This correspondence in results between different organisms, particularly in view of the small number of investigations involved at the level of the fish and bird, seems somewhat encouraging. It would appear that the length variables govern performance in animals at least as low on the phylogenetic scale as the fish, and moreover that the length variables govern performance in all three organisms in a highly similar fashion. Before this conclusion may be accepted, however, further research is needed, particularly along two lines. First, in the rat long N-length produces greater Rn than short N-length in much less than 400 trials (e.g., see Capaldi, 1958; Capaldi & Hart, 1962; and Weinstock, 1954, 1958). This may mean several things, among them that stimuli such as S^N and S^R are conditioned more rapidly in the rat than in the fish and perhaps in the bird. If this conjecture is correct, it may be expected that in order to produce the same sort of schedule of Reward × Acquisition Trial interaction in all three organisms, more acquisition training will be required at the level of fish and bird than at the level of the rat. It should also be indicated that the fish and bird investigations in which N length failed to increase Rn employed long intertrial intervals (24 hours). While it is not known whether long N-length will increase Rn in the fish and bird when a long intertrial interval is employed, the following comment seems in order. Even in the case of the rat, 50 acquisition trials is hardly sufficient to increase Rn as a function of N length (e.g., Capaldi, 1958). Thus, it is not unlikely that in those instances in which N-length failed to increase Rn in the fish and bird, the failure may have been more related to insufficient training than to intertrial interval. In the rat, of course, N-length will increase Rn both at short (e.g., Capaldi, 1958) and long (e.g., Weinstock, 1954, 1958) intertrial intervals. Thus, another possibility is that intertrial interval produces somewhat different effects in rat, fish, and bird. However, before this conclusion can be established it will be necessary to run N-length investigations on the fish and bird which employ considerable acquisition training as well as long intertrial intervals.

Dependent Variables

The theoretical curves presented earlier suggest that, in principle, the present hypothesis eventually should be capable of specifying the exact magnitude of the response on such and such a trial of extincton. Predictions of this sort are, of course, the ultimate aim of behavior theory. How seriously this capacity for specificity in prediction should be taken at the moment is difficult to say. Of course, any advance over the typical state of affairs, which is to predict that one group should show greater Rn than another (greater-than predictions), would be welcome. Quite possibly, the present hypothesis may be currently capable of delivering rate-of-extinction predictions of the general form, Group A should be faster than Group B in the early extinction trials but slower in the later ones. This proposition is advanced because, as indicated previously, the hypothesis has already enjoyed some success along this line (Capaldi, 1964; Gonzalez & Bitterman, 1964; Spivey, 1964). But rate predictions are far more sensitive to the specific conditions of an experiment than are greater-than predictions. Consider, for example, Boren's (1961) investigation which showed Rn increased as the size of the fixed-ratio (FR) increased. Boren provided "starter" reinforcements at a very low value of FR (2:1) before shifting some goups to higher training values of FR (5:1 to 20:1). According to the hypothesis, such pretraining should produce distortion in the shape of Boren's extinction curves. Under Boren's conditions, this distortion should have appeared mainly in the later extinction trials for smaller values of FR (e.g., 5:1) and mainly in the early extinction trials for higher values of FR (e.g., 20:1). In substance, when greater-than predictions are contemplated, the decision as to how many reinforcements to provide for this or that value of S^N is merely important; for rate predictions, the decision is absolutely crucial. Of course, purely practical considerations enter here. It is obviously rather difficult to get an animal to respond consistently at FR 20:1 without prior pretraining at, say, FR 2:1. In this connection, Boren notes that Rn does not seem to be further increased by FR beyond 20:1, specifically citing FR 50:1. Observations of this sort are helpful in determining when the limit of the stimulus modification process has been reached. Once it is reached, of course, further increases in N-length will not further increase Rn.

Other PR Variables: Magnitude of Reward and Intertrial Interval

Magnitude of reward. It has been shown (e.g., Hulse, 1958; Wagner, 1961) that under an irregular PR schedule Rn is an increasing function of reward magnitude. This finding may be understood within the framework of the present hypothesis by entertaining an assumption popularized and then later abandoned by Hull (1943, 1952) that the limit to which associative strength may grow is an increasing function of reward magnitude. This assumption has been more or less universally rejected in recent times in favor of incentive interpretations of the reward magnitude variable (e.g., Hull, 1952; Spence, 1956). Accordingly, it seems unlikely that learning theorists will be tempted to abandon the currently popular incentive viewpoint for the older, seemingly discredited associational one in the absence of a rather complete associational analysis of the magitude variable. An analysis of this type has been presented elsewhere (Capaldi, in press).

It is assumed that the greater the magnitude of reward under which a particular value of S^N is conditioned to the instrumental reaction the greater will be its associative capacity to evoke that reaction. Moreover, the greater the associative capacity of S^N to evoke its reaction, the greater will be the amount of generalized associative strength supplied to still higher values of S^N. It already has been shown that Rn is an increasing function of the associative capacity of the higher values of S^N to evoke the instrumental reaction.

The deduction that Rn will increase as magnitude of PR increases is applicable without reservation only to the case in which the different magnitude groups receive the same schedule of PR. Under such conditions all groups would have the same value of S^N conditioned to the reaction and each of the values would be conditioned with equal frequency; the determinate factor would thus be that the associative capacities of those S^Ns conditioned to the reaction would be greater, the greater the magnitude of reward. These theoretical conditions were met in the Hulse and the Wagner investigations since in both of them the irregular PR schedule employed was identical for the large reward and small reward groups.

The present hypothesis, being a sequential one, must expect that the effects of reward magnitude will depend to some extent on the schedule of PR. The general reason for such an expectation is, of course, that the associative capacity of S^N to evoke its reaction is a joint function of number of occurrences of N-length and magnitude of reward and under some conditions the former variable may well outweigh the latter. This may be illustrated at a relatively simple level by considering two PR groups which receive a small number of training trials. It was shown for small numbers of acquisition trials, it will be remembered, that Rn is primarily regulated by number of occurrences of N-length. Thus, a PR schedule employing small reward but many transitions from N trials to R trials could conceivably produce greater Rn than one employing large reward and a substantially fewer number of transitions from N trials to R trials. That is, S^N might possess a substantially greater capacity to evoke its reaction under the former schedule (more conditionings of S^N) than under the latter schedule. Experimental results bearing on this issue are not currently available; actually, it appears that no attempt has been made as yet to combine the sequence and magnitude variables in a single investigation. This is probably related to the fact that other PR hypotheses which have already dealt with magnitude of reward, for example, the frustration hypothesis (Amsel, 1958; Spence, 1960), have not as yet addressed themselves to the sequence variable.

Intertrial interval. The temporal characteristics of stimuli such as S^R and S^N are of systematic interest particularly since according to one interpretation (Hull, 1952; Sheffield, 1949) they are stimulus traces which dissipate relatively rapidly. However, and this does not seem to have been directly indicated previously, it is entirely possible to identify S^R and S^N as the response-produced stimuli of a mediating reaction (r_g-s_g). In this event, a concern with dissipation would be of considerably less relevance since the mediating reaction (r_g) is conditioned to external stimuli. Historically, a rather straightforward method has been employed to determine if S^R and S^N are in fact stimulus traces. As Sheffield (1949) has indicated, according to the trace conception, trial spacing per se should influence the type of experimental results obtained under PR since, by definition, stimulus traces are operative at shorter intertrial intervals (ITI) and inoperative at longer ones. It was on the basis of Weinstock's (1954, 1958) classical investigations that the interpretation of

S^R and S^N as stimulus traces was first rejected since PR was shown to increase Rn despite the employment of a 24-hour ITI. More recently, a number of investigations from our laboratory (Capaldi & Spivey, 1964a, 1964b; Capaldi & Stanley, 1963; Capaldi & Wargo, 1963) have supported Weinstock's conclusion in showing that a variety of PR variables produces approximately similar results under long and short ITIs.

However, rejecting the idea that S^R and S^N are stimulus traces is not to reject the idea that they regulate performance under PR. It seems relatively clear that if, as has been attempted in this paper, one can supply in terms of S^R and S^N a consistent explanation for massed and spaced trial PR data, then it is justifiable to assume that such stimuli are functional at long ITIs as well as short ones. In this event, as indicated, one plausible alternative is to view S^R and S^N as the response-produced stimuli of a mediating reaction. By doing so, however, we appear to commit ourselves to a somewhat different view of mediating reactions than has prevailed in animal learning since Hull's (1930) original paper on the topic. This may be clarified by examining two contrasting properties normally assigned to stimulus traces and mediating reactions.

The stimulus trace is considered to dissipate relatively rapidly, on the one hand, and to be more or less completely responsive to the current reward condition, on the other. The meaning of completely responsive is as follows: If the current trial is nonrewarded the trace on the succeeding trial will be that of nonreward. The mediating reaction is normally assigned somewhat opposite properties. It is, of course, considered to be long-lived. Moreover, it is considered to be only partially responsive to the current reward outcome. Thus, if r_g is strongly established, a single nonrewarded trial will reduce its strength through extinction only slighlty (e.g., Logan, Beier, & Kincaid, 1956).

According to the current PR hypothesis neither of these mechanisms is entirely satisfactory. What the present hypothesis requires is a stimulus which is like that produced by the mediating reaction in being long-lived but which is also like that produced by the trace mechanism in being more or less completely responsive to the current reward outcome. The former requirement, as already indicated, is dictated by the results of spaced-trial PR investigations, the latter requirement by the results of practically all PR investigations covered in this paper. In numerous places throughout this paper it was asumed that S^R and S^N were more or less completely responsive to the current reward outcome. In the absence of that assumption the present interpretation would not be feasible.

Considerations advanced in previous sections of this paper suggest that the notion of what might be called a completely responsive mediating reaction can be applied to extinction. Can it be applied to acquisition? The traditional or partly-responsive view of mediating reactions entertains two assumptions with acquisition: That r_g is formed slowly and that some number of trials are required to condition the stimulus it produces (s_g) to the instrumental reaction (e.g., Amsel, 1958). It appears possible to reject the former assumption and accept the latter one and still retain the capacity to deal with current acquisition data in mediation terms. That is, it may be assumed that the mediating reaction develops quickly but that the stimuli it produces are conditioned to the instrumental reaction rather slowly. This is to say in effect that expectancy reactions, that is,

r_g, are conditioned to their stimuli at a much more rapid rate than instrumental reactions are conditioned to their stimuli. At a common sense level, the present view suggests that the animal quickly learns about reward or nonreward or punishment in the experimental situation but that a number of additional trials are required before it learns what to do about its diffuse expectations. At a more formal level, the present view suggests that expectancy or mediating reactions and instrumental reactions are conditioned at different rates.

REFERENCES

AMSEL, A. The role of frustrative nonreward in non-continuous reward situations. *Psychological Bulletin,* 1958, **55,** 102–119.

BACON, W. E. Partial reinforcement extinction effect following different amounts of training. *Journal of Comparative and Physiological Psychology,* 1962, **55,** 998–1003.

BLOOM, J. M., & CAPALDI, E. J. The behavior of rats in relation to complex patterns of partial reinforcement. *Journal of Comparative and Physiological Psychology,* 1961, **54,** 261–265.

BOREN, J. J. Resistance to extinction as a function of the fixed ratio. *Journal of Experimental Psychology,* 1961, **61,** 304–308.

CAPALDI, E. J. The effect of different amounts of training on resistance to extinction of different patterns of partially reinforced responses. *Journal of Comparative and Physiological Psychology,* 1958, **51,** 367–371.

CAPALDI, E. J. Effect of N-length, number of different N-lengths, and number of reinforcements on resistance to extinction. *Journal of Experimental Psychology,* 1964, **68,** 230–239.

CAPALDI, E. J. A sequential hypothesis of the reward variables: Partial reward, delay of reward, and magnitude of reward. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation,* Vol. 1. New York: Academic Press, in press.

CAPALDI, E. J., & HART, D. Influence of a small number of partial reinforcement training trials on resistance to extinction. *Journal of Experimental Psychology,* 1962, **64,** 166–171.

CAPALDI, E. J., HART, D., & STANLEY, L. R. Effect of intertrial reinforcement on the aftereffect of nonreinforcement and resistance to extinction. *Journal of Experimental Psychology,* 1963, **65,** 70–74.

CAPALDI, E. J., & SENKO, M. G. Acquisition and transfer in partial reinforcement. *Journal of Experimental Psychology,* 1962, **63,** 155–159.

CAPALDI, E. J., & SPIVEY, J. E. Effect of goal-box similarity on the aftereffect of nonreinforcement and resistance to extinction. *Journal of Experimental Psychology,* 1963, **66,** 461–465.

CAPALDI, E. J., & SPIVEY, J. E. Intertrial reinforcement and aftereffects at 24-hour intervals. *Psychonomic Science,* 1964, **1,** 181–182. (a)

CAPALDI, E. J., & SPIVEY, J. E. Stimulus consequences of reinforcement and nonreinforcement: Stimulus traces or memory. *Psychonomic Science,* 1964, **1,** 403–404. (b)

CAPALDI, E. J., & STANLEY, L. R. Temporal properties of reinforcement aftereffects. *Journal of Experimental Psychology,* 1963, **65,** 169–175.

CAPALDI, E. J., & WARGO, P. Effect of transitions from nonreinforced to reinforced trials under spaced-trial conditions. *Journal of Experimental Psychology,* 1963, **65,** 318–319.

GONZALEZ, R. C., BEHREND, E. R., & BITTERMAN, M. E. Partial reinforcement in the fish: Experiments with spaced trials and partial delay. *American Journal of Psychology,* 1965, **78,** 198–207.

GONZALEZ, R. C., & BITTERMAN, M. E. Resistance to extinction in the rat as a function of percentage and distribution of reinforcement. *Journal of Comparative and Physiological Psychology,* 1964, **58,** 258–263.

GONZALEZ, R. C., ESKIN, R. M., & BITTERMAN, M. E. Further experiments on partial reinforcement in the fish. *American Journal of Psychology,* 1963, **76,** 366–375.

GONZALEZ, R. C., GRAF, V., & BITTERMAN, M. E. Resistance to extinction in the pigeon as a function of secondary reinforcement and pattern of partial reinforcement. *American Journal of Psychology,* 1965, **78,** 278–284.

GRANT, D. A., RIOPELLE, A. J., & HAKE, H. W. Resistance to extinction and the pattern of reinforcement: I. Alternation of reinforcement and the conditioned eyelid response. *Journal of Experimental Psychology*, 1950, **40**, 53–60.

GRANT, D. A., & SCHIPPER, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *Journal of Experimental Psychology*, 1952, **43**, 313–320.

GROSSLIGHT, J. H., HALL, J. F., & MURNIN, J. Patterning effects in partial reinforcement. *Journal of Experimental Psychology*, 1953, **46**, 103–106.

HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychological Review*, 1930, **37**, 511–525.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *A behavior system.* New Haven: Yale University Press, 1952.

HULSE, S. H., JR. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *Journal of Experimental Psychology*, 1958, **56**, 48–57.

JENSEN, G. D. Number and size of unrewarded blocks as determiners of resistance to extinction. *American Psychologist*, 1964, **19**, 534. (Abstract)

LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and varied reinforcement. *Journal of Experimental Psychology*, 1956, **52**, 65–70.

ROBERTS, W. A., BULLOCK, D. H., & BITTERMAN, M. E. Resistance to extinction in the pigeon after partially reinforced instrumental training under discrete trial conditions. *American Journal of Psychology*, 1963, **76**, 353–365.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *Journal of Experimental Psychology*, 1949, **39**, 511–526.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale University Press, 1956.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs: Prentice-Hall, 1960.

SPIVEY, J. E. Influence of small numbers of partial reinforcement training trials on resistance to extinction. Unpublished doctoral dissertation, University of Texas, 1964.

TYLER, D. W., WORTZ, E. C., & BITTERMAN, M. E. The effect of random and alternating partial reinforcement on resistance to extinction in the rat. *American Journal of Psychology*, 1953, **66**, 57–65.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *Journal of Experimental Psychology*, 1961, **62**, 234–252.

WAGNER, A. R., SIEGEL, S., THOMAS, E., & ELLISON, G. D. Reinforcement history and the extinction of a conditioned salivary response. *Journal of Comparative and Physiological Psychology*, 1964, **58**, 354–358.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *Journal of Comparative and Physiological Psychology*, 1954, **47**, 318–322.

WEINSTOCK, S. Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *Journal of Experimental Psychology*, 1958, **56**, 151–158.

(Received August 17, 1965)

Psychological Review
1966, Vol. 73, No. 5, 478–480

THEORETICAL NOTE

REINSTATEMENT[1]

BYRON A. CAMPBELL AND JULIAN JAYNES
Princeton University

Reinstatement is defined as periodic partial repetition of an experience such that it maintains the effects of that experience through time. This principle is demonstrated in a developmental study on the effects of early fear in rats, and is then discussed in relation to clinical and developmental theory.

In most of the phyla from arthropods to man early experience exerts a multiplicity of effects on adult behavior (Beach & Jaynes, 1954; Scott, 1962). Sometimes such effects are the simple persistence in adult behavior of habits formed early in life. In other instances it may be that early experience influences later behavior by structuring the individual's perceptual or response capacities. And in still others, there is a critical period of development during which some aspect of behavior, on which later behaviors depend, is learned and molded for life.

In this paper we suggest yet another mechanism. Although obvious and disarmingly simple, it yet seems to the authors of such neglected importance as to warrant this note and the coining of a term for it. By *reinstatement* we denote a small amount of partial practice or repetition of an experience over the developmental period which is enough to maintain an early learned response at a high level, but is not enough to produce any effect in animals which have not had the early experience. The following experiment is meant as a demonstration of this phenomenon in a commonly studied instance of learning.

METHOD

The subjects were 30 albino rats of the Wistar strain born and raised in the Princeton colony. They were divided into three groups of 10 each, with an equal number of males and females in each. The apparatus used was one commonly used in fear experiments (Campbell & Campbell, 1961). It consisted of two compartments separated by a door, a black one with a grid floor, and a white compartment with a solid metal floor. Shock could be administered to the grid of the black compartment. To two of the three groups an early fear-arousing experience was given in the black compartment. This consisted of placing the rat just after weaning, when approximately 25 days old, on the grid side of the apparatus with the door fixed so that the rat could not escape, then giving the rat 15 2-second 170-volt shocks on a 20-second variable interval schedule, taking aproximately 5 minutes, then removing the animal and placing him on the nonshock side for 5 minutes, and then repeating this entire procedure once. Thus each animal received a total of 30 shocks. At the end of this period the rat was removed and placed in a home cage. A third control group was run through this procedure without any shock being administered to the grid. During the next month a total of three shocks—the reinstatements—were given to one of the early experience groups and to the control group. These shocks were administered 7, 14, and 21 days after the original training session. The procedure was to administer, at some random number of seconds up to a minute after the animal was placed on the grid side of the apparatus, a single 2-second shock of the same intensity as before. The rat was then placed in the white compartment for an identical period of time and then returned to its home cage. On alternate weeks the animal was placed first on the nonshock side of the cage and then on the shock side, with half of the animals being placed on the shock side for the first reinstatement procedure and half on the safe side. Otherwise this procedure was precisely the same as the training procedure except that only 1 instead of 30 shocks was administered. The second pretrained group was given the same procedure except that no shock was administered. One week after the third reinstatement procedure, when the animals were 53 days of age, they were all tested for the effects of their early experience. This was done by placing them individually in the black compartment (where all of them had been shocked at one time or another) with the door removed so that the animal could run freely into the white compartment. The time spent in the white compartment over the ensuing hour was then recorded.

---

[1] This research was supported in part by Public Health Service Grant M-1562 from the National Institutes of Mental Health and by National Science Foundation Grant GB 2814.

RESULTS AND DISCUSSION

The results were unequivocal. As seen in Figure 1, the group that had received the early fearful experience followed by three 2-second shocks administered at weekly intervals, spent an increasing percentage of its time in the white compartment during the 1-hour test period, thus showing the effects of the early fearful experience with the black compartment. In contrast, the group that had had a similar early experience just after weaning, but no reinstatement of it in the intervening month, failed to show any significant fear of the black compartment, spending on the average all but about 10 minutes of the hour on that side. Similarly the group which had not had any early traumatic experience, but had received the three brief shocks over the month, failed to acquire any significant fear of the black compartment. The difference between the first group and the other two groups is, as it appears on the graph, highly reliable statistically ($p < .01$, Mann-Whitney U test).

There is nothing dramatically surprising about this finding. It is indeed what anyone thinking carefully about learning and practice would expect, namely, that there is some small amount of practice over certain time intervals which could maintain a previously learned response and yet not be enough to train naïve animals to perform that response. The possibility that this mechanism of reinstatement has wide and important applicability in the ontogeny of behavior in many vertebrate species seems beyond question.

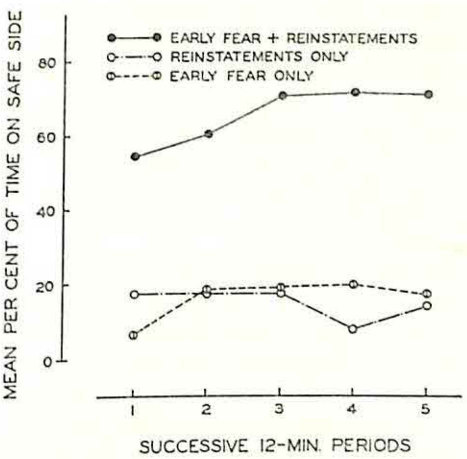

FIG. 1. The effect of reinstatement of early fear on later behavior.

In theoretical analyses of human growth and development traumatic events in infancy and childhood have long occupied a central, if controversial, role. In Freud's early analyses, traumatic events in childhood were considered a major cause of adult behavior disorders. With time, this view was gradually modified such that White writing in 1956 summed up current opinion by stating:

> Undoubtedly it is true that some adult neuroses have their origin in violently frightening events. . . . The theory has long since been abandoned, however, that all neuroses, or even a majority of neuroses, take their start from traumatic events [1956, p. 238].

The early trauma theory has inconsistencies with certain facts of memory and learning as well. First, on a mere phenomenological level, we know that memory becomes more and more dim the further back into our childhood we try to remember. Second, in rats, the earlier in life that a fearful experience is given the animal, the more likely it is to be forgotten in adulthood (Campbell & Campbell, 1962). Third, in chickens, the earlier in the critical period that the chick is imprinted, the more likely it is to be forgotten when the animal reaches the juvenile stages (Jaynes, 1957). This synchronized "sleep" patterns, it was hypothesized that this was due to the removal of the trigeminal and vestibular tonic barrage.

This view of sleep was prevalent until Moruzzi and Magoun (1949) showed that stimulation in parts of the midbrain RF of a cat put to sleep by barbiturates or chloralose yielded a desynchronized EEG. They interpreted their results as showing that the middle core of the midbrain contains a structure that diffusely arouses the organism and maintains it in an alert state. Their inference was based on the assumption that the desynchronized EEG represented an awake pattern.

If their interpretation were correct, then it should have been possible to eliminate any form of alertness in a chronic preparation by removing the structure. This is what Lindsley, Schreiner, Knowles, and Magoun (1950) did. They showed that cats with massive lesions in midbrain RF were asleep or comatose during the postoperative period. They also showed that lesions in lateral areas, involving the primary sensory pathways, did not produce these effects. From these studies and subsequent experiments (for reviews see French, 1960; Lindsley, 1960; Magoun, 1963; Samuels, 1959) the following picture of the brain mechanisms underlying sleep emerged. Bremer's *cerveau isolé* preparation showed a synchronized sleep pattern because of the severing of the RF from the cerebrum. The *encéphale isolé* preparation could show a desynchronized pattern in the absence of spinal inputs because the RF was still capable of bombarding the cortex. It was also pointed out that the activity of the RF was maintained by collaterals of primary sensory pathways into the RF (Starzl, Taylor, & Magoun, 1951). Thus sleep was considered a passive process, brought about by the inactivity of the RF, presumably caused by the reduction of sensory inputs into that structure.

RECENT DATA ON SLEEP

Aserinsky and Kleitman (1955) noted that during a night's sleep there were periods in which *S*s showed rapid eye movements (REM). These periods of REM occurred about every 60 minutes for 20–30 minutes; the duration of REM was longer in the later hours than in the beginning of sleep. Dement and Kleitman (1957b) showed that these periods of REM were related to periods in which *S* was most likely dreaming. They roused their *S*s both during periods of REM and periods where there were no rapid eye movements. They showed that the amount of dream recall was much greater during REM periods. These authors interpreted their results as demonstrating that dreaming occurred during REM. It should be pointed out, however, that dream recall is possible during other stages of sleep (Foulkes, 1964). Dement and Kleitman (1957a) also showed that during these REM periods the EEG showed a desynchronized pattern, similar to that seen in the waking human! Although this same result had been seen by Derbyshire, Rempel, Forbes, and Lambert (1936), interpretation was difficult since it was not clear whether these dreaming *S*s were sleeping or were, in fact, awake. If sleeping, why was their EEG record similar to that seen in the waking *S?* If awake, how were they dreaming?

Dement (1958) investigated these problems by studying the EEG patterns of the cat chronically implanted with recording electrodes. He kept the animal awake for 2 days, then fed it with warm milk and food, placed *S*, with electrode leads attached, in a quiet chamber, and waited for *S* to fall asleep. He found that there was a progressive synchronizing pattern as *S* went to sleep, until *S* showed a completely synchronized EEG. Following this, desynchronized EEG patterns suddenly appeared. The REM were seen, as well as twitching, and other small motor movements. This period was called Stage 1-REM. The progressively more synchronized stages of the EEG were called 2, 3, and 4. Following 1-REM, Stages 2, 3, and 4 successively appeared, and then the cycle began again. This cycle was repeated throughout the night. Dement (1958) also attempted to awaken *S* during the various stages of the EEG. He found that it was more difficult to do so during Stage 1-REM than during any other stage. Dement identified this Stage 1-REM as sleep, but was not certain whether it stood between wakefulness and sleep or represented a very deep stage of sleep. One could assume that this was a light stage of sleep, and that the reason it was difficult to arouse *S* was that *S* was distracted, attending to something else, most probably its dream. On the other hand, one could assume that this was a deeper stage of sleep than synchronized sleep, and that the reason *S* was not aroused was simply because of its depth of sleep. The man who has championed this latter view is Michel Jouvet. It is his work, and that of the Pisa school, along with these pioneering studies of Dement that have suggested revisions of certain notions concerning mechanisms of sleep and EEG. Before continuing, it will be wise to interrupt this history and briefly mention some problems in the interpretation of the EEG.

EEG IN SLEEP AND WAKEFULNESS

Until recently, the early EEG work of Rheinberger and Jasper (1937) remained more or less accepted. These investigators as well as others (Rossi & Zanchetti, 1957, p. 307, for references) had shown that the EEG exhibited several patterns which paralleled certain behavioral states. The low-voltage, high-frequency desynchronized pattern was often termed an aroused or awake EEG, while the high-voltage, low-frequency EEG pattern was often called the sleep EEG. It now appears that there are cases in which this parallelism does not stand up. The work of Dement, for example, suggests that it is possible to obtain a desynchronized awake EEG during sleep.

Jouvet (1962) has shown that the anticholinesterase eserine increases the desynchronized EEG pattern seen in the sleeping animal. It has also been possible to demonstrate a synchronized EEG pattern while *S* is fully awake. For example, Wikler (1952) and Bradley and Key (1958) have shown that *S*s injected with atropine show a synchronized EEG, but show behavioral alertness. Feldman and Waller (1962) have demonstrated the same result with selective brainstem lesions. In sum, desynchronized or low-voltage fast activity (LVF) may be associated with either sleep or wakefulness, and synchronized or high-voltage slow activity or slow waves (SW) may likewise be associated with either sleep or wakefulness.

These considerations do not eliminate the EEG as a useful tool in brain physiology, but they do make it quite clear that:

1. The use of the terms arousal pattern and sleep pattern in characterizing certain EEG outputs is sometimes incorrect. The substitution of the terms desynchronized, or low-voltage fast activity, and synchronized, or high-voltage slow activity or slow waves, seems preferable.
2. It is inadequate to use EEG records alone to infer level of consciousness without reference to the behavior

of *S* as well as other independent physiological measures. One of the major problems associated with the research on sleep is that of describing the state of awareness of *S*, since it is now clear that the EEG alone cannot be used, and that converging operations (Garner, Hake, & Ericksen, 1956) are necessary to demonstrate the presence or absence of consciousness. It would seem from these considerations that, rather than considering the EEG a useless neurophysiological tool, it would be more profitable to investigate the complexity which has been thus introduced in order to clarify those mechanisms underlying the synchronization and desynchronization of the EEG seen both during wakefulness and sleep.

WORK OF JOUVET

Originally published in French (Jouvet & Michel, 1958), Jouvet's work has been reported in English and French at several symposia (Jouvet, 1961a, 1961b, 1963); the most extensive summary of his work (Jouvet, 1962) has been reported in French. Essentially, Jouvet has continued the work of Dement (1958) and has elaborated the possible telencephalic (cortex, hippocampus, septal area, etc.), mesodiencephalic (thalamus, hypothalamus, midbrain RF), and rhombencephalic (pons, medulla) mechanisms that may be involved in the production of sleep. His experimental evidence and bold interpretations of data have made this area of research important in attacking many problems concerned with the organization of the central nervous system during sleep and waking.

Jouvet has worked mainly with cats, chronically implanted with electrodes in cortical, subcortical, and brainstem areas. He finds that there are basically two stages of sleep, which he calls "slow" and "fast" sleep. The former is sometimes called telencephalic and the latter rhombencephalic, to refer to the general anatomical area hypothesized to mediate the two stages. During the slow stage of sleep (S-SW), the EEG shows spindles (12–18/second) followed subsequently by SW (2–4 cycles/second) which appear in cortex, thalamus, and mesencephalic and pontine RF. During this time, the hippocampus shows high-voltage spikes. The electromygraphic potentials are present, and the midbrain RF stimulus necessary to wake *S* is 20–30% greater than during falling off to sleep. During sleep with low-voltage fast activity (S-LVF), or the paradoxical or rhombencephalic stage of sleep, there is low-voltage desynchronized activity, similar to that seen in the waking state. The hippocampus shows the arousal pattern of theta activity originally described by Green and Arduini (1954). In pontine RF, and more precisely in *n. reticularis pontis caudalis* (RPC), high-voltage waves (8/second) in the form of spindles are seen to occur. This 8/second activity is also seen in the interpeduncular nucleus and the posterior hypothalamus. While the electrographic pattern is in many respects like that of the waking *S*, the behavior of *S*, as well as certain physiological measures, indicate that *S* is asleep. Jouvet points out that at the outset of S-LVF there is a marked atonia; the head drops, and the electromyogram (EMG) of the nuchal muscles of the neck is flattened. As compared with the S-SW, respiratory movements during S-LVF are irregular, and the cardiac cycle may show either a distinct acceleration or a distinct deceleration. Eye movements are lateral and vertical, rapid and "explosive." There is twitching of the tail, vibrissae, and ears. Fast sleep always follows slow sleep and does not arise first.

Further comparisons of the slow and fast phases of sleep reveal that: (*a*) The threshold of arousal by either midbrain RF or peripheral sensory stimulation is appreciably raised during S-LVF; (*b*) evoked potentials are lowered during this period; and (*c*) S-LVF occupies 30–40%, while S-SW occupies 60–70% of the sleeping time of the cat.

Having described the major characteristics of sleep in both the slow and fast stages, Jouvet then proceeded to determine the anatomical loci responsible for these effects. Concerning the locus of slow sleep, Jouvet found that total neodecortication prevented the appearance of S-SW, but did not interfere with the occurrence of S-LVF. If there was any cortex remaining, then S-SW appeared. Concerning the locus of S-LVF, Jouvet found that sections in front of the pons prevented its appearance without preventing the appearance of S-SW, while sections behind the pons did not interfere with either the slow or fast phase of sleep. Reasoning that the structure mediating S-LVF was in pons, Jouvet made discrete electrolytic lesions in various cell masses within the pontine tegmentum. He found that lesions of RPC eliminated S-LVF, but did not affect S-SW. From these studies it was suggested that S-SW is brought about by cortical influences, and that sleep with desynchronized EEG, or rhombencephalic sleep (S-LVF), is triggered by RPC.

Jouvet also performed several lesion experiments in which he obtained evidence against the notion that the desynchronization seen during S-LVF is mediated through midbrain RF. Thus, lesions in dorsal tegmentum that abolished the classical EEG arousal response to nociceptive stimuli did not abolish the desynchronizing action seen during S-LVF. Jouvet has further shown that lesions in ventral tegmentum, posterior hypothalamus, interpeduncular nucleus, and septal region may reduce or eliminate S-LVF. All of these structures are included within the limbic midbrain circuit of Nauta (1958). Thus Jouvet has attempted to make a distinction between the desynchronization seen during arousal and the desynchronization seen during sleep. He has also stated that the EEG of these two states looks different to the "trained eye."

These findings may be related to those of Olds and Peretz (1960) and Glickman, Sroges, and Hunt (1964). The former found greater self-stimulation from midbrain RF electrodes located in the ventral region than in the dorsal region, and greater lower threshold arousal responses from more lateral and dorsal areas. Glickman et al. (1964) found that lesions in the ventral area increased activity markedly, while lesions in the more dorsal areas had no such effect. While it is not as yet clear what common element exists in S-LVF, self-stimulation, and exploratory behavior, the consistent differences between the dorsal and ventral RF with reference to these behaviors make the investigation of this problem worthwhile. As a possible working hypothesis one could assume that all three behaviors are related to some reinforcing consequence. That S-LVF may be reinforcing is suggested by the irritability of *S*s deprived of S-LVF (Dement, 1960). Along these lines, it is of interest to speculate whether septal rage (Brady & Nauta, 1953) may be at all ascribed to deprivation of S-LVF. These possibilities have not been investigated.

Two major conclusions can be drawn from the Jouvet experiments. First, it would seem fairly safe to say that S-LVF is a deeper stage of sleep than S-SW. Second, it appears that sleep is primarily an active process rather than a passive deafferentation. With respect to the first conclusion, Winters (1964) has suggested that S-LVF represents a stage of distraction rather than deep sleep and is thus intermediate between wakefulness and S-SW. Winters has pointed out that the major claim of Jouvet is the decreased responsiveness to peripheral stimuli during S-LVF. According to Winters, this decrease may be caused either by an increased excitability of neurons (occlusion) or a decreased excitability. Since Jouvet has not clarified this point, he cannot make inferences about the state of depth of sleep. It seems that Winters is assuming that depth of sleep is directly related to frequency of unit firing. There does not seem to be any evidence to warrant this assumption. Also, Jouvet stimulated midbrain RF directly and compared arousal thresholds at various stages of sleep. Winters does not mention these findings. Winters has also pointed out that the disappearance of the EMG is not sufficient evidence to define depth of sleep. Rossi and his co-workers (Candia, Favale, Guissani, & Rossi, 1962) have shown a marked drop in blood pressure during S-LVF. Giaquinto, Pompeiano, and Somogyi (1964) have shown a complete abolition of the heteronymous afferent reflex and the reduction of several other spinal reflexes during S-LVF. These effects do not occur as markedly in S-SW. In addition, petit mal epilepsy is depressed during S-LVF (Cadilhac, Vlahovitch, & Delange, 1965), suggesting that this stage of sleep may be deeper than S-SW. Thus, while Winters (1964) may be correct in saying that EMG disappearance *or* (if he knew of these data) blood-pressure drop *or* spinal reflex depression is insufficient evidence to demonstrate the depth of sleep, it would appear that all three lines of evidence taken together strongly favor the view that S-LVF is a deeper stage of sleep than S-SW. The work of Jouvet and his major conclusions have been supported by the reports of Hubel (1960), Rossi, Favale, Hara, Guissani, and Sacco (1961), and Moruzzi (1963a). See also the reviews by Cordeau (1962) and Moruzzi (1963b, 1964).

The second major conclusion is that the induction of sleep is probably an active process. The work of Hess (1957) and others (Rossi & Zanchetti, 1957, pp. 387–390) had shown that sleep could be induced by stimulation of particular brain loci. These studies were taken to be evidence for an active sleep system. It has remained a problem, however, to obtain this effect with frequencies greater than 15/second. It has therefore been argued that the induction of sleep is not related to the area stimulated but rather to the characteristic of the stimulus frequency. While this argument remains cogent for that body of evidence, it cannot refute the data of Jouvet which show the *absence* of sleep when certain structures are destroyed. Also, stimulation *at high frequencies* in pontine structures can bring about sleep (Jouvet, 1962; Rossi et al., 1961). It would appear quite safe to conclude that sleep is brought on by an active process. The next group of studies that we will consider will show some of the anatomical substrates underlying the synchronizing and desynchronizing patterns seen in the EEG.

WORK OF MORUZZI AND THE PISA GROUP

At about the same time as Jouvet published his work on the sleep patterns of the cat, Batini, Magni, Palestini, Rossi, and Zanchetti (1959), Batini, Moruzzi, Palestini, Rossi, and Zanchetti (1959), and Batini, Palestini, Rossi, and Zanchetti (1959) presented their work on the midpontine pretrigeminal preparation. A preliminary report was published in this country (Batini, Moruzzi, Palestini, Rossi, & Zanchetti, 1958). The major technique employed by these workers is similar to that of Bremer—namely, transection of the brain at various levels, and analysis and comparison of the EEG related to the various transections. They have found that transections in front of the trigeminal nerve give either persistent desynchronization or synchronization (spindles and SW) depending on the angle of the transection. Thus in the rostropontine pretrigeminal preparation, the EEG is predominantly synchronized, while in the midpontine pretrigeminal preparation the EEG is desynchronized. It may be asked whether the desynchronization seen in the midpontine preparation indicates arousal or an animal in S-LVF. Several experiments (e.g., Affanni, Marchiafava, & Zernicki, 1962a, 1962b; King & Marchiafava, 1963) show conclusively that *S* is awake.

These results have several important implications. First, it had previously been assumed that the

> . . . capacity to arouse is retained until the transection is made rostral to the trigeminal nucleus at which time the sleep-state of the *cerveau isolé* is induced. Clearly, therefore, somatic sensibility from the head is a more powerful contributor to tonic reticular excitation than are all other input systems [French, 1960, p. 1288].

The work of Batini, Magni, Palestini, Rossi, and Zanchetti (1959), Batini, Moruzzi, Palestini, Rossi, and Zanchetti (1959), and Batini, Palestini, Rossi, and Zanchetti (1959) has shown that it is possible to achieve persistent patterns of desynchronization with the deafferentation of the trigeminal nerve. This does not exclude the possibility that the trigeminal contributes to the activation of RF, but trigeminal input is certainly not necessary for EEG desynchronization patterns. In fact, Batini, Palestini, Rossi, and Zanchetti (1959) have shown that while desynchronization in the midpontine pretrigeminal preparation disappears when the other two sensory inputs (optic, olfactory) are severed, even in this case desynchronization patterns begin to appear after 2 or 3 days. These authors have concluded that in the "midpontine pretrigeminal preparation EEG activation patterns are independent of the sensory inflow [p. 32]."

The above findings suggest revisions for reticular theorists since it has been repeatedly emphasized that EEG activation is brought about by sensory inputs and is maintained by corticofugal fibers to RF. But Morillo and Baylor (1963) have found that the medial lemniscus, spinothalamic tract, and spinocerebellar tract do not give off collaterals to the midbrain RF. Inasmuch as these major sensory pathways have, at times, been claimed to do so, much work must be done to clarify exactly what inputs reach the RF and to what extent such inputs innervate this structure.

It has been mentioned that one implication of the work of Batini, Magni, Palestini, Rossi, and Zanchetti (1959) was the demonstration of the noncritical nature of the trigeminal in the desynchronization of the EEG. The question of what structure maintained the persistent desynchronization still remains. Since a section at the midbrain pontine border (rostropontine) led to synchrony, and a lesion at the middle of the pons (midpontine) led to persistent desynchrony, it would seem that the tissue between these two cuts was necessary for the persistent desynchronization. If this were so, it is not yet clear whether this tissue represents (*a*) an extension of the midbrain RF, or (*b*) a separate structure independent, to some extent, of the midbrain RF. If the former, then the obtained result could be explained on the basis of mass action; that is, a certain amount of midbrain and pontine RF is necessary to desynchronize cortex. Removal of some of this structure will decrease its ability to desynchronize. If the latter, it may be possible that the pontine structure is more critical than midbrain RF in maintaining wakefulness. The pontine structure which has been tentatively identified is *n. reticularis pontis oralis.*

Before continuing it is necessary to point out a persistent problem with lesion research, particularly experiments in which entire sections of the brain are prevented from influencing one another.

DENERVATION SUPERSENSITIVITY

Psychologists and physiologists have generally ignored the problem set forth by Stavraky (1961). When a brain structure is removed by an experimental or vascular lesions, the structures which were innervated by the removed structure will change their properties. The change in property will usually be in the form of an increased chemical sensitivity to other incoming stimuli (either naturally or experimentally applied). This finding has been demonstrated for peripheral structures (Cannon & Rosenbleuth, 1949), but it has not been conclusively shown that denervation supersensitivity exists in the central nervous system. Sharpless and Halpern (1962) reported that after several weeks of little spontaneous activity chronically isolated cortex sometimes showed EEG patterns which "resembled that recorded from normal cortex in waking animal [p. 246]." These results, while only suggestive, make possible another interpretation of the Batini, Palestini, Rossi, and Zanchetti (1959) finding that cortex can still show a desynchronized pattern without sensory inputs. Perhaps the desynchronization seen 2 days postoperatively in the midpontine pretrigeminal totally deafferented preparation is not due to the intrinsic activity of the RF alone, but rather is potentiated by a generalized supersensitivity resulting from denervation of all sensory inputs to the brain. Whatever the state of affairs, it is quite clear that the importance of denervation supersensitivity has not been adequately realized (but see Harvey & Lints, 1965), perhaps because it can make the interpretation of any neurophysiological experiment in which a lesion is made extremely difficult. There is little data on this issue. Perhaps subsequent experiments will take this phenomenon into account.

SYNCHRONIZING MECHANISMS

In the Batini experiments conducted in 1959, it may not only be fruitful to ask what mechanisms are directly responsible for the *presence* of desynchronization, but it may also be worthwhile to ask what mechanism might underlie the *persistence* of desynchronization. If it were assumed that there was a synchronizing mechanism in caudal pons that inhibited rostral pontine and midbrain RF, then one could explain the results of the Pisa group by assuming that in the midpontine pretrigeminal preparation, the synchronizing mechanism is cut off from the activating structures in pons and midbrain. Thus, persistent desynchronization would be obtained by release from inhibition. This view is supported by different workers in different laboratories using a variety of techniques. The main conclusion that emerges is that there is a powerful inhibitory influence being exerted by a synchronizing mechanism located in caudal pons and perhaps rostral medulla. Also, the presence of such a mechanism reinforces the suggestion made earlier that sleep is brought about by active inhibitory mechanisms and essentially is not occurring passively.

The first major suggestion that there was a synchronizing influence on cortical activity was made by Cordeau and Mancia (1958, 1959). They studied the EEG of cats that sustained midbrain hemisections. These authors found that in the awake cat there was greater EEG synchronization in cortex on the same (ipsilateral) side as the midbrain section as compared to the contralateral side. During the first few postoperative days, when *S* was put to sleep by pentobarbital or went to sleep naturally, synchronization or sleep spindles appeared first on the side ipsilateral to the lesion. After 6–8 postoperative days, the asymmetry of the EEG in the two hemispheres seen in the awake *S* was gone. When *S* now went to sleep, the sleep spindles first appeared on the *opposite* side (contralateral) of the midbrain hemisection. If this same experiment was performed with the addition of a pretrigeminal midpontine section, all aspects of the above result were repeated except for failure of the sleep spindles to appear on the contralateral side of the midbrain hemisection after 6–8 postoperative days. Cordeau and Mancia suggested that these results favored the notion of a synchronizing sleep system which exerts its effect behind the midpontine section. The reason that the switch occurred in the locus of onset of spindles in the midbrain-hemisectioned *S* is likely the result of a compensatory action of the brain tissue subserving the synchronizing function. In the midpontine pretrigeminal preparation, on the other hand, the compensatory switch is not seen postoperatively, since the mechanism producing these spindles was severed from the rest of the brain. In order to explain why spindles are produced at all in the pretrigeminal preparation, it is assumed that the thalamic spindle pacemaker (Morrison & Dempsey, 1943) does operate in the absence of the lower brainstem synchronizing influence.

Magni, Moruzzi, Rossi, and Zanchetti (1959) have demonstrated arousal reactions by the application of the barbiturate thiopental. These workers were able to limit the distribution of the drug to the lower brainstem by ligation procedures which did not interrupt the vertebral or carotid circulation. Under these conditions the application of thiopental produced a desynchronized response in the EEG. A similar application of the same drug, but allowing its distribution throughout the brain, produced the expected EEG synchronization. These authors interpreted their results as supporting the view that a synchronizing structure is located within the lower pons and rostral medulla. Application of thiopental depressed the activity of this structure, thereby releasing the pontine and midbrain RF from its inhibitory influence, and bringing about a desynchronized response. That the lower brainstem was, in fact, depressed is revealed by the absence of the corneal reflex. These results confirm those of Cordeau and Mancia (1958, 1959) and support the notion that there exists a synchronizing mechanism in the lower brainstem.

Hugelin, Bonvallet, and their coworkers (Bonvallet & Allen, 1963; Bonvallet & Bloch, 1961; Dell, Bonvallet, & Hugelin, 1961) have shown that there are both cortical and lower brainstem effects acting on the midbrain RF of Moruzzi and Magoun (1949). Briefly summarizing their results, any procedure which depressed cortical tone, for example, decortication, cooling, small doses of chloralose, hyperventilation, increased the excitability of RF as evidenced by the greater reticular facilitation of the masseteric myotatic reflex. From their findings they postulated a negative-feedback system in which excitation of RF of the midbrain leads to cortical arousal which then dampens reticular activity. These results may be contrasted with the facilitatory positive-feedback system which French, Hernández-Peón, and Livingston (1955) described. These latter workers showed that stimulation in cortex produced reticular arousal and suggested that the cortex is important in the maintenance of arousal. Since it is quite possible that both inhibition and facilitation of RF by cortical structures may be present, the conflicting results of the Parisian and Californian workers may be only apparent.

Of more importance to the present discussion is the review of the Bonvallet and Bloch (1960) study by Dell et al. (1961). These authors showed that the interruption of the pathways between the RF of the midbrain and bulbar structures increased the intensity and duration of cortical arousal produced by afferent nerve stimulation. Also, stimulation in RF produced a much greater arousal in the prebulbar sectioned animal than in the *encéphale isolé* preparation. These data likewise lend support to the notion of a synchronizing system in lower pons and rostral medulla.

It is not clear, however, how this synchronization is brought about. Dell et al. (1961) consider two possibilities. Either this synchronizing mechanism exerts its effect on the thalamic synchronizing system or it exerts its effect by deactivating RF. Dell et al. favor the latter view, but it is not at all necessary to reject the former view in the process, since it is quite likely that one result of reticular deactivation is the partial release of the synchronizing system of the thalamus from the overpowering effects of reticular stimulation (Purpura & Shofer, 1963).

Bonvallet and Allen (1963) have performed a series of experiments to determine the location of the synchronizing structure and whether its effect is mediated by deactivating RF or by activating synchronizing structures. These authors have shown that discrete lesions of the cephalic portion of the *n. tractus solitarius* are followed by a general instability of reticular activity in sympathetic, parasympathetic, motor, and cortical facilitation functions. These findings suggest a loss of control of RF function by the elimination of a homeostatic mechanism in medulla and strongly implicate the *n. tractus solitarius* as the critical synchronizing mechanism. These authors also showed by other lesion experiments that the effects obtained could not be duplicated by destruction of the classical bulbar inhibitory RF described by Magoun and Rhines (1946). Thus, while the area most clearly implicated in the latter effect has been the ventromedial RF, Bonvallet and Allen (1963) found that the "synchronizing" structure was located more laterally. They have localized the effect to the *n. tractus solitarius* at the level of the dorsal motor nucleus of the tenth nerve and to those cells lying between this area and the floor of the fourth ventricle. They have also shown that the locus of the synchronizing effect is distinct from the nearby bulbar region regulating blood pressure, but may be involved with afferent fibers from the cardiovascular system.

Another example of inhibitory influence from lower structures will be given, primarily because the conclusions are somewhat similar to those reported above, but the neurophysiological technique used is quite different.

Demetrescu and Demetrescu (1962) studied the effects of reticular stimulation on potentials evoked in visual cortex by electrical stimulation of the lateral geniculate. They found that stimulation in dorsal areas of RF enhanced visual evoked potentials, but stimulation in ventral pontine RF had no effect. If lesions were then made in the central gray of the midbrain, stimulation of ventral RF inhibited the thalamo-cortical evoked potential. This study suggests that there is an inhibitory system located in ventral pons which exerts an effect that is opposite in sign to that shown by the dorsal and lateral midbrain tegmentum.

A recent report by Rossi, Minobe, and Candia (1963) has dealt with the problem of identifying the site of origin of the ascending influence responsible for the EEG patterns of deep sleep. They have performed brain hemisection experiments similar to those of Cordeau and Mancia (1958, 1959). Rossi et al. (1963) have compared EEG patterns in cats hemisectioned at various levels of the brainstem during slow and fast sleep. Of particular interest to our discussion on synchronizing mechanisms is the finding that there was no asymmetry in the EEG with hemisections at the caudal pons and rostral medulla, suggesting that the synchronizing system is probably located at the pons-medulla boundary region. These authors showed that hemisection of the midbrain produced an increase in S-SW and a decrease in S-LVF. When the hemisection was placed behind *n. reticularis pontis oralis*, there was an increase in S-LVF and a decrease in S-SW. When the lesion was placed at the level of RPC, there was a decrease of S-LVF and an increase in S-SW. According to Rossi et al. (1963), then, both slow and fast sleep were apparently controlled by a system located within the pons, and it thus appeared that:

> . . . the integrity of the rostral part of the pons is essential for the maintenance of the waking state, and the middle and caudal parts of the pons play a very important role in the occurrence of sleep. A different rostro-caudal distribution of the arousing (rostral pons) and sleep-inducing (middle and caudal pons, and perhaps medulla oblongata) neurons seem proved [p. 487].

These authors stated that there appears to be an active and passive origin of slow sleep and an active origin of fast sleep. In the former case, the passive origin refers to the deafferentation of midbrain RF at the first stages of sleep, while the active origin of slow sleep refers to the inhibitory action of pontine synchronizing structures on pontine and perhaps midbrain RF.

There are several difficulties in interpreting the Rossi et al. (1963) study. The one difficulty that appears most outstanding is the possibility that sections at certain levels of the brainstem may be eliminating both synchronizing and desynchronizing systems or synchronizing systems and systems concerned with the mechanisms of S-LVF. In such cases, it would seem difficult to make any definite statements concerning functional localization. Subsequent studies should be performed using discrete stereotaxic lesions.

Summarizing this section on synchronizing systems, it would appear certain that there are important structures in the caudal pons that play a role in synchronizing the EEG during sleep. Further studies should clarify the relation between this brainstem synchronizing system and the synchronizing system found in thalamus (Morrison & Dempsey, 1943).

SINGLE-UNIT ACTIVITY

Several experiments have been reported in which correlations have been made between the state of consciousness of the organism and brain activity recorded presumably from individual neurons by means of microelectrodes .5–10 microns in diameter at the uninsulated tip. From these experiments it has been shown that, in general, single-unit activity is at a high level during wakefulness, at a low level during S-SW, and at a high or higher level during S-LVF than during wakefulness. This result has been shown with neuronal activity derived from cortex (Evarts, 1962; Murata & Kameda, 1963), pyramidal system (Arduini, Barlucchi, & Strata, 1963; Evarts, 1964), and from vestibular nuclei (Bizzi, Pompeiano, & Somogyi, 1964). Increases in single-unit activity in RF during S-LVF have, in fact, been demonstrated by Huttenlocher (1961) and Brooks and Bizzi (1963).

It appears quite remarkable that the individual elements of the central nervous system are as active during wakefulness as during S-LVF. One possibility suggested by Evarts (1964) is that the patterning of neuronal activity is different during wakefulness than during S-LVF. Thus, frequency of unit firing would be similar during wakefulness and during S-LVF, but the distribution of activity (e.g., interspike interval, variance) would be different. In this regard, the analysis of thalamic-unit activity by Werner and Mountcastle (1963) may perhaps be fruitfully applied to the comparison between neuronal firing during wakefulness and neuronal firing during S-LVF.

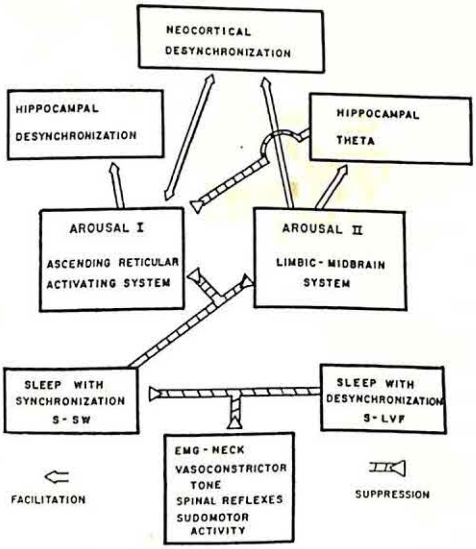
NEOCORTICAL DESYNCHRONIZATION
HIPPOCAMPAL DESYNCHRONIZATION
HIPPOCAMPAL THETA
AROUSAL I
ASCENDING RETICULAR ACTIVATING SYSTEM
AROUSAL II
LIMBIC-MIDBRAIN SYSTEM
SLEEP WITH SYNCHRONIZATION S-SW
SLEEP WITH DESYNCHRONIZATION S-LVF
EMG-NECK
VASOCONSTRICTOR TONE
SPINAL REFLEXES
SUDOMOTOR ACTIVITY
FACILITATION
SUPPRESSION

FIG. 1. Possible brainstem organization involved in sleep. (White arrows denote facilitation, striped arrows, suppression.)

ORGANIZATION OF RETICULAR SLEEP MECHANISMS

An attempt will be made in this section to bring together the results discussed up to this point into a schema of brainstem function during sleep. The basic system is summarized in Figure 1. Certain features of this schema should be pointed out. First, it is assumed that *there are two anatomically distinct desynchronizing systems,* which we shall call Arousal System I and Arousal System II. System I is closely related to the classical arousal system originally described by Moruzzi and Magoun (1949) and is generally thought of as *the* arousal system. System II is part of the limbic-midbrain system described by Nauta (1958) in the cat. Although arousal as a unitary phenomenon has been used fruitfully in neuropsychological theory (Hebb, 1955), the present data as well as other findings (e.g., Feldman & Waller, 1962) appear to necessitate the postulation of two arousal systems. Since neocortical desynchronization does not differentiate between the activity of Systems I and II, it is important to determine whether some other measure can differentiate between desynchrony produced by System I and desynchrony produced by System II. It appears that electrical activity in hippocampus may serve such a need. It will be recalled that Green and Arduini (1954) showed that stimulation of RF produced 4–7 cycle/second waves in hippocampus. These authors identified this hippocampal theta rhythm with arousal. Recently, however, it has been shown that stimulation in midbrain RF can also produce LVF (desynchronization) in hippocampus (Stumpf, 1965). This result suggests the possibility that neocortical EEG desynchronization may indeed be divided into two states: one identified with hippocampal desynchronization and another identified with hippocampal theta. The former, in the present schema, is derived from reticular-activating System I, the latter, from limbic-midbrain System II (Kawamura, Nakamura, & Tokizane, 1961).

A second feature of the schema is that *hippocampal theta tends to dampen the activity of System I.* This view is primarily based on the work of Grastyan (1959), who has interpreted his results to mean that hippocampal activity represents an inhibition of the orienting reflex to novel stimuli. A third feature of the schema is that *a positive-feedback loop exists between System I and neocortex.* Such a process has been suggested by French et al. (1955). Based on the above view that hippocampal theta dampens the activity of System I, and that neocortex and System I form a positive-feedback system, the following hypothesis may be stated: At low or intermediate levels of activity, System II will predominate over System I since the suppressing effect of hippocampal theta will tend to diminish the positive-feedback effect between System I and neocortex. At high levels of activation, System I will predominate since the feedback system will be in too great a state of activity to be suppressed by hippocampal theta.

From the present review of the sleep literature, there appear to be two critical structures in brainstem involved in the sleep process. The first structure is the synchronizing mechanism which has been tentatively identified with the rostral aspect of the *n. tractus solitarius.* The second structure is the S-LVF mechanism which has been shown by Jouvet (1962) and others (see Hobson, 1965) to be associated with *n. reticularis pontis caudalis.* With respect to the synchronizing mechanism, the present schema assumes that both Arousal I and Arousal II are inhibited by this system. According to the present view, the midpontine section of Batini, Magni, Palestini, Rossi, and Zanchetti (1959) separates the synchronizing system from Arousal I and Arousal II, thus causing a release from the inhibition of the synchronizing system. If it were assumed that *n. reticularis pontis oralis* represents a critical component of the positive-feedback system, then one might understand how a rostropontine section might yield significantly less neocortical desynchronization.

It remains finally to account for the relation of the S-LVF mechanism (RPC) with the system thus far described. The assumption that satisfies most of the demands of the data is that RPC inhibits the synchronizing mechanism. In such a case, inhibition of the synchronizing system would remove the inhibition from Arousal I and Arousal II and thus permit the latter's activity. This activity would then appear as desynchronization in the neocortical EEG and theta rhythm in hippocampus. If we assume that RPC directly or indirectly inhibits neck-muscle EMG, vasoconstrictor tone, sudomotor activity, and spinal reflexes, then the present schema has achieved a representative description of what has been called "paradoxical sleep," that is, S-LVF.

Two comments concerning possible cortical relations with RPC and possible relation of this model to behavior in the waking state will be made. Presumably, if telencephalon is critical for the onset of S-SW (Jouvet, 1962), and if RPC is important in S-LVF, it would be useful to know what relation, if any, exists between them. Newman and Wolstencroft (1959) have recorded short-latency responses of 1 millisecond in medulla resulting from stimulation of the orbital gyrus in frontal lobe. This result was also obtained by these authors using the strychnine neuronography method. Earlier, Sachs, Brendler, and Fulton (1949) had noted strychnine spikes in medial forebrain bundle (MFB), and Nauta (1962) has recently traced fibers from the posterior orbital area to MFB. Auer (1956) has also noted terminal degeneration in MFB following frontal cortex ablation. No degeneration was seen in the ventromedial nucleus of the hypothalamus or in mammillary nuclei. These studies suggest that the orbital gyrus has a strong relation to the limbic-midbrain system described by Nauta (1958). It will be recalled that this circuit was implicated strongly by Jouvet as important in S-LVF.

The second comment concerns possible use of the schema in the waking state. It should be pointed out, first, that the process of sleep seems organized on a hierarchical basis. Considering four states, aroused, relaxed, S-SW, and S-LVF, transitions from one stage to another by proper stimulation can be demonstrated. Thus, several workers have been able to produce relaxation and then S-SW with low-frequency stimulation (Pompeiano & Swett, 1962; see Rossi & Zanchetti, 1957, for review) or with chemical stimulation (Hernández-Peón, Chavez-Ibarra, Morgane & Timo-Iara, 1963). The reverse operations have also been shown by Lissak, Karmos, and Grastyan (1963); they were able to produce sleep with slow waves following S-LVF by stimulation in hippocampus. The tentative conclusion may be made that in the intact and properly functioning organism, there is a successive reduction in the locus of control of the conscious state. As Jouvet has pointed out, S-LVF is always preceded by S-SW. Thus it would appear that the difference between the sleep and waking state may be a function of the system which is in primary control and in the overall *pattern* of nervous activity in central and peripheral structures, rather than total overall activity alone. This means that any single measure will probably never describe the state of consciousness of the organism. It would appear that at least two measures will be required.

Having discussed the problems with inferring consciousness from a single measure, it now may be fruitful to consider the possibility that the synchronizing and desynchronizing mechanisms seen during sleep may be operating during wakefulness. For example, it has been shown by Clemente, Sterman, and Wyrwicka (1964) that immediately following reinforcement the EEG of the cat is synchronized. This pattern of synchrony in the awake animal may have its derivation from either thalamic or brainstem synchronizing structures or both. We have no data on this point. Since it has been shown that both synchrony and desynchrony appear in both sleep and wakefulness, perhaps our understanding of the neural mechanisms of the sleep process will allow for insights into the neural mechanisms of waking behavior as well.

CERTAIN CHANGING VIEWS OF RF FUNCTION

Consideration of the neural mechanisms underlying sleep suggests that there may be a necessity to modify current views of RF function.

1. The brainstem organization important in sleep and wakefulness is extremely complex. The proposed schema represents only the merest suggestion of the complexity that is involved. The notion that sleep results from midbrain RF inactivity is at best a simplification, and at worst, incorrect (Moruzzi, 1963b). Whatever the correct organization of the system may be, it is clear that the view of the single, unified RF that is either active (wakefulness) or inactive (sleep) is inadequate to account for the current data on sleep.

2. LVF (desynchronization) seen on an EEG may represent high arousal, moderate arousal, or deep sleep. Clearly this desynchronization should not be termed an "arousal pattern" in the absence of other correlated measures.

3. While a major emphasis has been placed on the *midbrain* RF as the critical structure in mediating arousal, the work of Moruzzi, Rossi, and their coworkers suggests that the *pontine* RF plays at least as important a role in sleep and wakefulness as does midbrain RF.

4. Feldman and Waller (1962) have shown that certain lesions of the brainstem caused an animal to be asleep or drowsy while producing a desynchronized EEG; other lesions caused a synchronized EEG pattern, but the animal was quite awake. It would appear that present concepts of RF function are inadequate to account for such a result. The present schema might be able to do so if it were assumed that one lesion was in Arousal I while the other lesion was in Arousal II. Similar problems in interpretation occur with the work of Wikler (1952), who showed that atropine causes synchronized EEG patterns but leaves the animal awake. In this case, also, it may be possible that atropine differentially affects the two arousal systems so that wakefulness is maintained by one, while synchrony is brought about by the atropine-caused depression of the other.

5. Finally, it may be pointed out that several traditional experimental findings and interpretations have been questioned by recent research. Thus, Schlag and Chaillet (1962) have shown that the diffuse thalamic projection system does not *directly* cause neocortical desynchronization, but rather activates cortex by a descending route through brainstem RF. Baust, Niemeczyk, and Vieth (1963) and Baust and Niemeczyk (1963, 1964) have shown quite conclusively that the adrenergic arousal reported by Rothballer (1956) was a secondary effect caused by a primary response of an increase in blood pressure. Although by no means definite, Morillo and Baylor (1963) found no evidence that medial lemniscus projects into the brainstem RF. Such a result is opposite to the findings of Starzl et al. (1951). More work is required to resolve this issue.

REFERENCES

AFFANNI, J. M., MARCHIAFAVA, P. L., & ZERNICKI, B. Conditioning in the midpontine pretrigeminal cat. *Archives italiennes de biologie*, 1962, **100**, 305-310. (a)

AFFANNI, J. M., MARCHIAFAVA, P. L., & ZERNICKI, B. Orienting reflex in the midpontine pretrigeminal cat. *Archives italiennes de biologie*, 1962, **100**, 287–304 (b)

ARDUINI, A., BARLUCCHI, G., & STRATA, P. Pyramidal activity during sleep and wakefulness. *Archives italiennes de biologie*, 1963, **101**, 530–544.

ASERINSKY, E., & KLEITMAN, N. Two types of ocular motility occurring in sleep.

Journal of Applied Psychology, 1955, **8,** 1–10.

AUER, J. Terminal degeneration in the diencephalan after ablation of frontal cortex in the cat. *Journal of Anatomy,* London, 1956, **90,** 30–41.

BATINI, C., MAGNI, F., PALESTINI, M., ROSSI, G. F., & ZANCHETTI, A. Neural mechanism underlying the enduring EEG and behavioral activation in the midpontine pretrigeminal cat. *Archives italiennes de biologie,* 1959, **97,** 15–25.

BATINI, C., MORUZZI, G., PALESTINI, M., ROSSI, G. F., & ZANCHETTI, A. Persistent patterns of wakefulness in the pretrigeminal midpontine preparation. *Science,* 1958, **128,** 30–32.

BATINI, C., MORUZZI, G., PALESTINI, M., ROSSI, G. F., & ZANCHETTI, A. Effects of complete pontine transections on the sleep-wakefulness rhythm: The midpontine pretrigeminal preparation. *Archives italiennes de biologie,* 1959, **97,** 1–12.

BATINI, C., PALESTINI, M., ROSSI, G. F., & ZANCHETTI, A. EEG activation patterns in the midpontine pretrigeminal cat following sensory deafferentation. *Archives italiennes de biologie,* 1959, **97,** 26–32.

BAUST, W., & NIEMCZYK, H. Studies on the adrenalin sensitive component of the maesencephalic reticular formation. *Journal of Neurophysiology,* 1963, **26,** 692–704.

BAUST, W., & NIEMCZYK, H. Further studies on the action of adrenergic drugs on cortical activity. *Electroencephalography and Clinical Neurophysiology,* 1964, **17,** 261–271.

BAUST, W., NIEMCZYK, H., & VIETH, I. The action of blood pressure on the ascending reticular activating system with special reference to adrenalin-induced EEG arousal. *Electroencephalography and Clinical Neurophysiology,* 1963, **15,** 63–72.

BIZZI, E., POMPEIANO, O., & SOMOGYI, I. Spontaneous activity of single vestibular neurons of unrestrained cats during sleep and wakefulness. *Archives italiennes de biologie,* 1964, **102,** 308–330.

BONVALLET, M., & ALLEN, M. B., JR. Prolonged spontaneous and evoked reticular activation following discrete bulbar lesions. *Electroencephalography and Clinical Neurophysiology,* 1963, **15,** 969–988.

BONVALLET, M., & BLOCH, V. Le controle bulbaire des activations corticales et sa mise en feu. *Compte rendu de la Société de biologie,* Paris, 1960, **154,** 1428–1431.

BONVALLET, M., & BLOCH, V. Bulbar control of cortical arousal. *Science,* 1961, **133,** 1133–1134.

BRADLEY, P. B., & KEY, B. J. The effect of drugs on arousal responses produced by electrical stimulation of the reticular formation of the brain. *Electroencephalography and Clinical Neurophysiology,* 1958, **10,** 97–110.

BRADY, J. V., & NAUTA, W. J. H. Subcortical mechanisms in emotional behavior: Affective changes following septal forebrain lesions in the albino rat. *Journal of Comparative and Physiological Psychology,* 1953, **46,** 339–346.

BREMER, F. Cerveau isolé et physiologie du sommeil. *Compte rendu de la Société de biologie,* Paris, 1935, **118,** 1235–1241.

BREMER, F. The neurophysiological problem of sleep. In E. D. Adrian, F. Bremer, H. H. Jasper, & J. F. Delafresnaye (Eds.), *Brain mechanisms and consciousness.* Oxford: Blackwell, 1954. Pp. 137–162.

BROOKS, D. C., & BIZZI, E. Brain stem electrical activity during deep sleep. *Archives italiennes de biologie,* 1963, **101,** 648–665.

CADILHAC, J., VLAHOVITCH, B., & DELANGE, M. Considerations on the changes in epileptic discharges during the phase of eye movements. *Electroencephalography and Clinical Neurophysiology,* 1965, **18,** 96.

CANDIA, O., FAVALE, E., GUISSANI, A., & ROSSI, G. F. Blood pressure during natural sleep induced by electrical stimulation of the brain stem reticular formation. *Archives italiennes de biologie,* 1962, **100,** 216–233.

CANNON, W. B., & ROSENBLEUTH, A. *The supersensitivity of denervated structures.* New York: MacMillan, 1949.

CLEMENTE, C. D., STERMAN, M. B., & WYRWICKA, W. Post-reinforcement EEG synchronization during alimentary behavior. *Electroencephalography and Clinical Neurophysiology,* 1964, **16,** 355–365.

CORDEAU, J. P. Functional organization of the brain stem reticular formation in relation to sleep and wakefulness. *Revue canadienne de biologie,* 1962, **21,** 113–125.

CORDEAU, J. P., & MANCIA, M. Effect of unilateral chronic lesions of the midbrain on the electrocortical activity of the cat. *Archives italiennes de biologie,* 1958, **96,** 374–399.

CORDEAU, J. P., & MANCIA, M. Evidence for the existence of an EEG synchronization mechanism originating in the lower brain stem. *Electroencephalography and Clinical Neurophysiology,* 1959, **11,** 551–564.

Dell, P., Bonvallet, M., & Hugelin, A. In G. E. W. Wolstenholme & M. O'Connor (Eds.), *The nature of sleep.* London: Churchill, 1961. Pp. 86–107.

Dement, W. The occurrence of low voltage, fast electroencephalogram patterns during behavioral sleep in the cat. *Electroencephalography and Clinical Neurophysiology,* 1958, **10**, 291–296.

Dement, W. The effect of dream deprivation. *Science,* 1960, **132**, 1420–1422.

Dement, W., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements. *Electroencephalography and Clinical Neurophysiology,* 1957, **9**, 673–690. (a)

Dement, W., & Kleitman, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346. (b)

Demetrescu, M., & Demetrescu, M. Ascending inhibition and activation from the lower brain stem: The influence of pontine reticular stimulation of thalamo-cortical evoked potentials in cat. *Electroencephalography and Clinical Neurophysiology,* 1962, **14**, 602–620.

Derbyshire, A. J., Rempel, B., Forbes, A., & Lambert, E. F. The effects of anesthetics on action potentials in the cerebral cortex of the cat. *American Journal of Physiology,* 1936, **116**, 577–596.

Evarts, E. V. Activity of neurons in visual cortex of cat during sleep with low voltage fast EEG activity. *Journal of Neurophysiology,* 1962, **25**, 812–816.

Evarts, E. V. Temporal patterns of discharge of pyramidal tract neurons during sleep and waking in the monkey. *Journal of Neurophysiology,* 1964, **27**, 152–171.

Feldman, S. M., & Waller, H. J. Dissociation of electrocortical activation and behavioral arousal. *Nature,* London, 1962, **196**, 1320–1322.

Foulkes, D. Theories of dream formation and recent studies of sleep consciousness. *Psychological Bulletin,* 1964, **62**, 236–247.

French, J. D. The reticular formation. In J. Field (Ed.), *Handbook of physiology.* Vol. 2. Washington, D. C.: American Physiological Society, 1960. Pp. 1281–1305.

French, J. D., Hernández-Peón, R., & Livingston, R. B. Projections from cortex to cephalic brain stem (reticular formation) in monkey. *Journal of Neurophysiology,* 1955, **18**, 74–95.

Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychological Review,* 1956, **63**, 149–159.

Giaquinto, S., Pompeiano, O., & Somogyi, I. Supraspinal modulation of heteronymous monosynaptic and of polysynaptic reflexes during natural sleep and wakefulness. *Archives italiennes de biologie,* 1964, **102**, 245–281.

Glickman, S. E., Sroges, R. W., & Hunt, J. Brain lesions and locomotor exploration in the albino rat. *Journal of Comparative and Physiological Psychology,* 1964, **58**, 93–100.

Grastyan, E. The hippocampus and higher nervous activity. In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Josiah Macy, Jr. Foundation, 1959. Pp. 119–205.

Green, J. D., & Arduini, A. A. Hippocampal electrical activity in arousal. *Journal of Neurophysiology,* 1954, **13**, 533–557.

Harvey, J. A., & Lints, C. E. Lesions in the medial forebrain bundle: Delayed effects on sensitivity to electric shock. *Science,* 1965, **148**, 250–252.

Hebb, D. O. Drives and the C.N.S. (Conceptual Nervous System). *Psychological Review,* 1955, **62**, 243–254.

Hernández-Peón, R., Chavez-Ibarra, G., Morgane, P. J., & Timo-Iara, C. Limbic cholinergic pathways involved in sleep and emotional behavior. *Experimental Neurology,* 1963, **8**, 93–111.

Hess, W. R. *The functional organization of the diencephalon.* (Ed. by J. R. Hughes.) New York: Grune & Stratton, 1957.

Hobson, J. A. The effects of chronic brain-stem lesions on cortical and muscular activity during sleep and waking in the cat. *Electroencephalography and Clinical Neurophysiology,* 1965, **19**, 41–62.

Hubel, D. H. Electrocorticograms in cats during natural sleep. *Archives italiennes de biologie,* 1960, **98**, 171–181.

Huttenlocher, P. R. Evoked and spontaneous activity in single units of medial brain stem during natural sleep and waking. *Journal of Neurophysiology,* 1961, **24**, 451–468.

Jouvet, M. Recherches sur les mechanismes neurophysiologiques du sommeil et de l'apprentissage négatif. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning.* Oxford: Blackwell, 1961. Pp. 445–479. (a)

JOUVET, M. Telencephalic and rhombencephalic sleep in the cat. In G. E. W. Wolstenholme & M. O'Connor (Eds.), *The nature of sleep.* Boston: Little, Brown, 1961. Pp. 188–208 (b)

JOUVET, M. Recherches sur les structures nerveuses et les mechanismes responsables des différentes phases du sommeil physiologique. *Archives italiennes de biologie,* 1962, **100**, 125–106.

JOUVET, M. The rhombencephalic phase of sleep. In G. Moruzzi, A. Fessard, & H. H. Jasper (Eds.), *Brain mechanisms.* New York: Elsevier, 1963. Pp. 405–424.

JOUVET, M., & MICHEL, F. Recherches sur l'activité électrique cérébrale au cours du sommeil. *Compte rendu de la Société de biologie,* Paris, 1958, **152**, 1167–1170.

KAWAMURA, H., NAKAMURA, Y., & TOKIZANE, T. Effect of acute brain stem lesions on the electrical activities of the limbic system and neocortex. *Japanese Journal of Physiology,* 1961, **11**, 564–575.

KING, F. A., & MARCHIAFAVA, P. L. Ocular movements in the midpontine pretrigeminal preparation. *Archives italiennes de biologie,* 1963, **101**, 149–160.

LINDSLEY, D. B. Attention, consciousness, sleep and wakefulness. In J. Field (Ed.), *Handbook of physiology.* Vol. 3. Washington, D. C.: American Physiological Society, 1960. Pp. 1553–1593.

LINDSLEY, D. B., SCHREINER, L. H., KNOWLES, W. B., & MAGOUN, H. W. Behavioral and EEG changes following chronic brain stem lesions in the cat. *Electroencephalography and Clinical Neurophysiology,* 1950, **2**, 483–498.

LISSAK, K., KARMOS, G., & GRASTYAN, E. A study of the so-called "paradoxical phase" of sleep in cats. In G. Moruzzi, A. Fessard, & H. H. Jasper (Eds.), *Brain mechanisms.* New York: Elsevier, 1963. Pp. 424–428.

MAGNI, F., MORUZZI, G., ROSSI, G. F., & ZANCHETTI, A. EEG arousal following inactivation of the lower brain stem by selective injection of barbituate into the vertebral circulation. *Archives italiennes de biologie,* 1959, **97**, 33–46.

MAGOUN, H. W. *The waking brain.* (2nd ed.) Springfield, Ill.: Charles C Thomas, 1963.

MAGOUN, H. W., & RHINES, R. An inhibitory mechanism in the bulbar reticular formation. *Journal of Neurophysiology,* 1946, **9**, 165–171.

MORGAN, C. T. *Physiological psychology.* New York: McGraw-Hill, 1965.

MORILLO, A., & BAYLOR, D. Electrophysiological investigation of lemniscal and paralemniscal input to the midbrain reticular formation. *Electroencephalography and Clinical Neurophysiology,* 1963, **15**, 455–464.

MORRISON, R. S., & DEMPSEY, E. W. Mechanism of thalamo-cortical augmentation and repetition. *American Journal of Physiology,* 1943, **138**, 279–308.

MORUZZI, G. Discussion of sleep mechanisms. In G. Moruzzi, A. Fessard, & H. H. Jasper (Eds.), *Brain mechanisms.* New York: Elsevier, 1963. Pp. 438–439. (a)

MORUZZI, G. The physiology of sleep. *Endeavour,* 1963, **22**, 31–36. (b)

MORUZZI, G. Reticular influences on the EEG. *Electroencephalography and Clinical Neurophysiology,* 1964, **16**, 2–17.

MORUZZI, G., & MAGOUN, H. W. Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology,* 1949, **1**, 455–473.

MURATA, K., & KAMEDA, K. The activity of single cortical neurons of unrestrained cats during sleep and wakefulness. *Archives italiennes de biologie,* 1963, **101**, 306–331.

NAUTA, W. J. H. Hippocampal projections and related neural pathways to the midbrain in the cat. *Brain,* 1958, **81**, 319–340.

NAUTA, W. J. H. Neural associations of the amygdaloid complex in the monkey. *Brain,* 1962, **85**, 505–520.

NEWMAN, P. P., & WOLSTENCROFT, J. H. Medullary responses to stimulation of orbital cortex. *Journal of Neurophysiology,* 1959, **22**, 516–523.

OLDS, J., & PERETZ, B. A motivational analysis of the reticular activating system. *Electroencephalography and Clinical Neurophysiology,* 1960, **12**, 445–454.

POMPEIANO, O., & SWETT, J. E. EEG and behavioral manifestations of sleep induced by cutaneous nerve stimulation in normal cat. *Archives italiennes de biologie,* 1962, **100**, 311–342.

PURPURA, D. P., & SHOFER, R. J. Intracellular recording from thalamic neurons during reticulocortical activation. *Journal of Neurophysiology,* 1963, **26**, 494–505.

RHEINBERGER, M. B., & JASPER, H. H. Electrical activity of the cerebral cortex of the unanesthetized cat. *American Journal of Physiology,* 1937, **119**, 186–196.

ROSSI, G. F., FAVALE, E., HARA, T., GUISSANI, A., & SACCO, G. Researches on the nervous mechanisms underlying deep sleep in the cat. *Archives italiennes de biologie,* 1961, **99,** 270–292.

ROSSI, G. F., MINOBE, K., & CANDIA, O. An experimental study of the hypnogenic mechanisms of the brain stem. *Archives italiennes de biologie,* 1963, **101,** 470–492.

ROSSI, G. F., & ZANCHETTI, A. The brain stem reticular formation: Anatomy and physiology. *Archives italiennes de biologie,* 1957, **95,** 199–435.

ROTHBALLER, A. B. Studies of the adrenaline-sensitive component of the reticular activating system. *Electroencephalography and Clinical Neurophysiology,* 1956, **8,** 603–621.

RUCH, T. C., PATTON, H. D., WOODBURY, J. W., & TOWE, A. L. *Neurophysiology.* Philadelphia: Saunders, 1965.

SACHS, E., JR., BRENDLER, S. J., & FULTON, J. F. The orbital gyri. *Brain,* 1949, **72,** 227–240.

SAMUELS, I. Reticular mechanisms and behavior. *Psychological Bulletin,* 1959, **56,** 1–25.

SCHLAG, J. D., & CHAILLET, F. Thalamic mechanisms involved in cortical desynchronization and recruiting responses. *Electroencephalography and Clinical Neurophysiology,* 1962, **15,** 39–62.

SHARPLESS, S. K., & HALPERN, L. M. The electrical excitability of chronically isolated cortex studied by means of permanently implanted electrodes. *Electroencephalography and Clinical Neurophysiology,* 1962, **14,** 244–255.

STARZL, T. E., TAYLOR, C. W., & MAGOUN, H. W. Ascending conduction in reticular activating system with special reference to the diencephelon. *Journal of Neurophysiology,* 1951, **14,** 461–477.

STAVRAKY, G. W. *Supersensitivity following lesions of the nervous system.* Toronto: University of Toronto Press, 1961.

STUMPF, C. The fast component in the electrical activity of the rabbit's hippocampus. *Electroencephalography and Clinical Neurophysiology,* 1965, **18,** 477–486.

WERNER, G., & MOUNTCASTLE, V. B. The variability of central neural activity in a sensory system, and its implication for the central reflection of sensory events. *Journal of Neurophysiology,* 1963, **26,** 958–977.

WIKLER, A. Pharmacologic dissociation of behavior and EEG "sleep patterns" in dogs: Morphine, n-allylmorphine and atropine. *Proceedings of the Society of Experimental Biology and Medicine,* 1952, **79,** 261–265.

WINTERS, W. D. Comparison of the average cortical and subcortical evoked response to clicks during various stages of wakefulness, slow wave sleep and rhombencephalic sleep. *Electroencephalography and Clinical Neurophysiology,* 1964, **17,** 234–243.

(Received September 23, 1965)

Psychological Review
1966, Vol. 73, No. 6, 500–526

RAPID EYE MOVEMENT SLEEP AND CORTICAL HOMEOSTASIS[1]

HARMON S. EPHRON
Department of Psychiatry, New York Medical College

AND

PATRICIA CARRINGTON
Teachers College, Columbia University

Findings indicate that under most conditions nonrapid eye movement sleep (NREMS) represents an indispensable condition for the onset of rapid eye movement sleep (REMS). The implications of this sequential relationship are explored. It is conceptualized that deepening NREM sleep involves a progressive loss of cerebral vigilance which, nevertheless, must somehow be maintained within adaptively appropriate limits. A homeostatic interplay between sleep phases is then postulated. This involves 2 complementary tendencies; 1 tendency toward deepening NREM sleep, and organismic rest, which, when it reaches a preset level, triggers the release of another tendency toward REMS and organismic activation. REMS is seen as acting to increase cortical "tonus" through a process of "endogenous afferentation."

Recent research has produced a rapidly expanding body of knowledge concerning two qualitatively distinct types of sleep: rapid eye movement sleep (REMS) and the quiescent sleep stages referred to collectively as nonrapid eye movement sleep (NREMS). In spite of much that has been learned in the past several years about the characteristics of REMS and NREMS, however, the physiological function of REMS is not yet known, and to date little attention has been paid to possible implications of the sequential relationship between these two forms of sleep; or to the significance of their markedly different—often contrasting—natures, for an understanding of the function of REMS.

Should REMS be looked upon as a regressive state which atavistically intrudes during sleep, or does it serve some current role in the economy of the organism? Does REMS operate on its own time clock independent of NREMS (which precedes it under most normal conditions) or may the cyclic alternation of these two sleep forms reflect some homeostatic process at work?

In this paper we propose to explore the relationship between the sleep phases, giving particular attention to evidence which suggests the presence of a homeostatic process underlying the NREMS-REMS cycle. Following a brief description of research and an examination of some other views on the function of REMS, we will outline a theory of the function of this sleep state. Implications of this concept for further research will be suggested.

RESEARCH AND THEORY

The NREMS-REMS Cycle

Since several comprehensive reviews of the literature are available (Dement, 1964; Dement, 1965a; Fisher, 1965; Hartmann, 1965; Snyder, 1965a,

---

[1] A shorter version of this paper was read before the Association for the Psychophysiological Study of Sleep, Washington, D. C., March 26–28, 1965. The authors would like to express appreciation to Howard Roffwarg and Frederick Snyder for their critical reading of an earlier draft of this manuscript and to Wilse Webb for his helpful comments during preparation of the final version.

1965b) we will recount only a few findings. By now it is well known that in 1953 Aserinsky and Kleitman noticed that periods of highly active (REM) sleep appeared to alternate with periods of quiet (NREM) sleep in a regular, cyclical manner, and that the activated sleep phase was accompanied by rapid conjugate eye movements which occurred in bursts (Aserinsky & Kleitman, 1953). It has subsequently been determined that this rapid eye movement state is present in human beings of all ages (Aserinsky & Kleitman, 1953, 1955) as well as in all mammalian forms studied to date (Klein, 1963; Snyder, 1965a, 1965b). In a typical night's sleep (in the human adult) REMS begins a little more than an hour after sleep onset, recurs roughly every 90 minutes throughout the night, and is accompanied by marked changes in electrical recordings, corresponding to a cyclic variation in sleep stages as measured by the electroencephalogram (EEG) (Dement & Kleitman, 1957a, 1957b).

Commonly, five stages of EEG sleep are identified. At sleep onset, a brief period of Stage 1 occurs, characterized by low voltage fast waves. This soon gives way to Stage 2, which is marked by the appearance of synchronous bursts of 14-cycles per second activity known as "sleep spindles" and/or by "K Complexes." Sleep then usually proceeds to Stages 3 and 4 ("deep" NREMS) which are characterized by slow, high voltage potentials. After some time in these "deeper" stages, there is a return to Stage 1, which is now for the first time accompanied by rapid eye movements and other signs of intense activation which have given rise to its separate designation as Stage 1-REM. Thus in normal subjects (*S*s), REMS regularly follows NREMS. This sequence of sleep stages is then repeated throughout the night with minor variations (Dement & Kleitman, 1957a).

Many distinct physiological changes are seen during the rapid eye movement state. While these are too numerous to list in their entirety, they include acceleration and irregularity of heart-rate (Snyder, Hobson, Morrison, & Goldfrank, 1964); increase in and variability of systolic blood pressure (Snyder et al., 1964); shallow, rapid and irregular breathing (Snyder et al., 1964); full or partial penile erections (in primates) (Fisher, Gross, & Zuch, 1965; Gross, Weitzman, Fisher, & Byrne, 1966; Karacan & Snyder, 1966); and increased brain temperature in cats (Rechtschaffen, Cornwell, & Zimmerman, 1965) and rabbits (Kawamura & Sawyer, 1965). Neural changes suggesting marked activation of sensory receiving and motor areas of the brain also occur and will be discussed later. Paradoxically, in contrast to numerous indications of increased neural and metabolic activity, resting muscle tone from head and neck virtually disappears during REMS (Berger, 1961; Jouvet & Michel, 1960a) and there is a sharp attenuation of spinal reflexes at this time (Hodes & Dement, 1964).

Some findings indicate that REMS may represent a necessary physiological state since both human *S*s (Dement, 1960) and animals (Dement, 1965b; Jouvet, 1965a; Khazan & Sawyer, 1963) systematically deprived of REMS (by being awakened shortly after onset of the rapid eye movements) show increased REMS on subsequent nights, time spent in this sleep stage on "recovery nights" (or number of awakenings required to maintain REMS deprivation) increasing in nearly linear fashion with the number of nights of such deprivation.

REMS appears related to dreaming in humans in such manner that dreams are reported in approximately 80% of awakenings from it (Dement & Kleitman, 1957a, 1957b). While investigators have obtained reports of mental content including dreams or "dreamlike" experiences from *all* phases of sleep (Foulkes, 1962; Goodenough, Shapiro, Holden, & Steinschriber, 1959; Kamiya, 1961; Orlinsky, 1962; Rechtschaffen, Verdone, & Wheaton, 1963), if one were to concentrate solely on the type of dream which includes highly organized, imagery-laden plots or stories with relatively more reference to strong emotion, rapid eye movement periods appear predictive of this sort of dream.

Theories Concerning the Function of REMS

Most theories to date on the function of REMS have been expressed briefly by authors reporting research. Speculations include suggestions that REMS may act as a "rhythmic energy-feeding mechanism" serving to maintain the organism through the long fast of sleep (Mandell & Mandell, 1965); REMS may be the expression of a periodic function of storing information at the molecular level and thus play an essential role in memory (Jouvet, 1965a); and REMS in prematures, newborns, and infants may correspond to a unique and ontogenetically primitive state of vigilance (Petre-Quadans, in press).

Although these and similar theories are provocative, the tentative form of their presentation has prompted us to confine our discussion to the more detailed hypothesis recently put forward by Roffwarg, Muzio, and Dement (1966). These authors have suggested that a functional consequence of REMS may be to assist structural maturation of the central nervous system (CNS) through supplying great quantities of endogenous stimulation to the higher centers at a time (in utero and during early postnatal life) when brain growth is maximum. They present arguments in support of the position that maturation is often dependent on function and point out that the extremely high percentage of REM sleep seen at birth falls off with increasing age and that, interestingly, it is at this time that the CNS is receiving correspondingly longer periods of waking stimulation.

The developmental theory of REMS proposed by Roffwarg, Muzio, and Dement has much to commend it and offers, in our opinion, the best explanation so far advanced for the much higher percentage of REMS seen in early life. This theory appears incomplete, however, in that it does not explain the continued occurrence of REMS throughout the life cycle. This limitation is noted by the authors, who suggest that REMS may fulfill some as yet unknown function of CNS stimulation even after maturation is completed. In the present paper, an attempt will be made to explore the function of REMS within the perspective of the organism's total life span.

The Concept of Cortical Tonus

Head (1923) used the term "vigilance" to describe a state of "high-grade neurophysiological efficiency" which varies with the state of the organism, the level of vigilance being gauged by the extent to which, at any given moment, the activity of a particular portion of the CNS exhibits signs of integration and purposive adaptation. Since this concept seems directly applicable to the study of sleep, some of its details may be noteworthy.

"Vigilance" is to be distinguished from "excitability." While the latter is inversely proportional to the strength of the excitation needed to produce any recognizable sign of activity (that is, to threshold value), the main factor in vigilance is the fundamentally *divergent nature* of the response. This would vary according to whether the organism is (at the time of stimulus presentation) in a high or low state of vigilance. A high state of vigilance is said to tend to evoke more or less high-grade, adapted responses.

While Head's concept appears promising for a comparative study of the various sleep stages, the term "vigilance" is commonly used to refer to a state of watchfulness with respect to danger or hazard. It is also employed in an experimental sense, to denote the degree of responsiveness to external stimuli found in the sleeping organism. It would therefore be desirable to find a different term to denote variations in the general level of efficiency or "tone" of the CNS.

In 1938, Bremer referred to "cortical tonus," a condition characterized by spontaneous electrical activity of the higher brain centers which was said to be maintained by corticopetal (afferent) impulses and thus to be automatically lowered with surgical or functional deafferentation. While it is not quite clear whether Bremer intended the term "cortical tonus" to refer to neurophysiological efficiency as well as neuronal activation, the term appears a useful one, and will be clarified for use in the present discussion.

We employ "cortical tonus" to denote a level of tonic CNS excitation suitable for efficient processing of information, limiting the term "vigilance" to denote responsiveness to external stimuli only. While cortical tonus and vigilance so defined will often covary, they need not always do so. The organism could be said to be in a state of heightened cortical tonus, for example, when it is responding to stimuli which are entirely endogenous in nature, as, for example, when organismically-generated afferent impulses or memory traces command attention. Since attentiveness to "endogenous afferentation" might, in this case, even impede responsiveness to external stimuli, heightened cortical tonus would then result in lowered vigilance.

While a certain level of neuronal excitation must be present for adequate cortical tonus, supernormal CNS excitement such as occurs when body temperature is experimentally raised or an excess of oxygen is administered can lead to hyperirritability, loss of concentration and other symptoms of *lowered* CNS efficiency (Barcroft, 1938). Accordingly, to establish the presence of adequate cortical tonus, we should find evidence of a heightened capacity for effective processing of information, as well as indications of neuronal activation. In evaluating evidence for a homeostatic theory of REMS it therefore seems desirable to consider both behavioral and neurophysiological data. In certain instances, however, data will be available in only the one or the other of these areas, thus limiting discussion to a single type of evidence.

Cortical Homeostasis

At the commencement of sleep, with descent into NREMS, cortical functioning is profoundly altered as the desynchronized neuronal activity of waking life gives way to synchrony. As sleep deepens there is a progressive loss of vigilance, although the organism does not lose the capacity to respond to urgent stimuli (Coleman, Gray, & Watanabe, 1959; Fischgold, Schwarz, & Dreyfus-Brisac, 1959; Oswald, Taylor, & Treisman, 1960). Since vigilance is not entirely lost in normal sleep (presumably because of the high survival value attached to the capacity to respond while asleep to certain vital challenges) it could be said that sleep is contained within *adaptively appropriate limits.*

While regulation of depths of sleep may seem self-evident, this is possibly only because we take for granted certain mechanisms which control the level of cerebral excitation so effectively that a sufficient degree of cortical tonus is regularly maintained. It might be useful to explore this question and search for means by which such tonus may be insured in sleep. As Cannon (1929) points out:

> in an open system such as our bodies represent, compounded of unstable material and subjected to continually disturbing conditions, constancy is in itself evidence that agencies are acting, or ready to act to maintain this constancy [p. 424]. . . . When a factor is known which can shift a homeostatic state in one direction it is reasonable to look for automatic control of that factor or for a factor or factors having an opposing effect [p. 426].

Cannon was primarily concerned with mechanisms (neural, endocrine, and others) which perform autonomic regulations necessary for normal bodily function and health. More recently, this classical concept of homeostasis has been extended by Davis (1950) and Gellhorn (1953) to include regulation of the CNS. Davis (1950) suggests that the regulation of the physiological state which normally varies diurnally between sleep and waking (and which extends, when homeostasis is inadequate, to coma at the one extreme and to convulsions at the other) may be looked upon as representing a *homeostasis of cerebral excitability.* Such a process would imply a dual function of the CNS: one part or subsystem performing the tasks usually associated with the nervous system, and the other (probably anatomically coextensive with the first) devoted to maintaining the proper level of CNS excitability or "tone." Davis does not see this regulatory mechanism within the nervous system as comprising the reticulo-cortical system or any other specific tracts, but suggests that a general system plus several subsystems is most likely involved. Since negative feedback would be required, his theory implies the equivalent of intracortical sense organs, themselves sensitive to the level of excitability of cortical neurons. We will later have occasion to refer to the possibility of such an afferent system within the cortex.

Gellhorn (1953) similarly proposes that there are present in the brain physiological adjustment reactions which operate to maintain the degree of cortical excitation within fairly narrow limits. Reasoning from a somewhat different point of view, he cites evidence suggesting that with excitation of the cortex, an antagonistic mechanism is automatically called into action which *reduces* excessive activation; and that, conversely, any condition leading to a *decrease* in cortical functions automatically calls forth a *release* of the brainstem from inhibition, which in turn leads to a return of cortical reactivity. Where surgical or functional decortication occurs, the state of the organism may pass from coma to convulsions as soon as the release of brainstem functions has progressed far enough, provided no narcotics are used which depress the brainstem. In light of this process, which Gellhorn terms *cortical homeostasis*, the convulsions are said to represent an ultimate adjustment mechanism (although of a pathological form) to counteract the coma through discharges from the brainstem and particularly from the hypothalamus. Sometimes the improvement of cortical functions seen as a result of a spontaneous or induced convulsive seizure is marked and *S* appears normal for some time, offering, according to Gellhorn, an impressive illustration of the homeostatic value of the seizure under certain conditions.

Applying the concept of cortical homeostasis[2] to the problem of the maintenance of cerebral excitation during sleep, we have conceptualized a homeostatic interplay between NREMS (characterized by quiescence and loss of cortical tonus), and REMS (characterized by excitation and heightened cortical tonus). To explore the implications of this concept, we will commence with discussion of the NREM phase.

The NREM Phase

Functional Deafferentation in NREMS

Oswald (1962) has described sleep as "a recurrent, healthy condition of inertia and unresponsiveness" and Magoun (1963) has pointed out that the sleeping state is characterized by the absence of perception. Similarly, Bremer (1935, 1938), Kleitman (1963), and Broughton (1966) have referred to sleep as causing "functional deafferentation."

The degree of deafferentation in sleep appears to vary with sleep stage. Experimentation indicates that as NREMS proceeds from Stage 1 toward Stages 3 and 4, various forms of response may disappear consecutively, with the greatest impairment being found in Stage 4. With progression into the deeper NREM sleep stages, reaction times for simple manual responses have been shown to increase (Coleman et al., 1959; Fischgold et al., 1959); EEG responses to auditory stimuli to decrease (Zung & Wilson, 1961); and discriminative capacity to become markedly lowered, or to disappear entirely as *S* reaches Stage 4 (Oswald et al., 1960).

Despite decreased responsiveness in the "deeper" sleep stages, however, evidence suggests that considerable deafferentation is experienced even in Stages 1 (descending) and 2. When Rechtschaffen and Foulkes (1965) taped open the eyes of sleeping *S*s and illuminated various objects held directly in front of them, they observed that in two of the three *S*s studied "relative functional blindness" seemed to set in even before the onset of Stage 1. These investigators found no evidence of incorporation of stimuli into subsequent verbal reports from *any* of the sleep stages, and had the impression that functional blindness commenced during the phase of EEG alpha rhythm and slow conjugate eye movements which usually precede unambiguous sleep. This suggests that functional deafferentation, at least in the visual sphere, may be virtually total at sleep onset.

Similarly, while *S*s are less responsive to auditory stimuli during sleep Stages 3–4 than they are during Stage 2, male *S*s studied by Zung and Wilson (1961) were unresponsive to auditory stimuli in at least half of the trials administered during Stage 2, with female *S*s unresponsive to such stimuli in about 25–30% of the trials during this stage (Zung & Wilson, 1966). It seems that the differences which exist between Stages 2, 3, and 4 sleep with respect to the degree of deafferentation experienced may be considerably less impressive than the dramatic contrast between sleep and the waking state with respect to this variable. We will therefore refer to NREMS, collectively, as "deafferenting sleep."

A paradoxical aspect of loss of vigilance in sleep is that certain evoked potentials at the cortical level may be *augmented* as the individual sinks into the deeper NREM sleep stages (Favale, Loeb, Manfredi, & Sacco, 1965;

---

[2] As used in this paper, this term incorporates Davis' concept of "homeostasis of cerebral excitability," as well.

Herz, 1965). Cortical excitability in slow-wave sleep thus appears similar to that seen under general anesthesia, where evoked potentials are sometimes displayed with unusual clarity. Such phenomena tend to occur, in fact, only in states where the brain has undergone profound modification of functions so that discriminatory responses to external stimuli are sharply reduced or absent.

Because these clearly evoked responses occur at times of relative unconsciousness, it has been assumed that the arousal of the primary neural volley (i.e., the arrival of an "incoming signal") at cortex from a stimulus is not sufficient to produce *perception*. It seems, rather, that more (or different) neuronal activity must be present to evoke *awareness* of the stimulus—apparently an evoked response must be "processed" by the brain in order to be effective as communication (Dustman & Beck, 1965). Certain evoked cortical potentials in sleep might thus be likened to a telephone ringing in a room with no one picking it up. The "signal" may enter the cortex in a crude fashion during sleep, but precise information is not conveyed because structures capable of processing this information appear to be unresponsive at this time.

In this paper such incompleted afferentation during sleep is referred to as "functional deafferentation." The result of such "deafferentation," as Magoun (1963) has pointed out, is a profound alteration in the interaction between the organism and the external world. This in turn could be expected to markedly lessen the organism's adaptive capacities.

Mental Operations during Sleep

The question of the effectiveness of cortical functioning during sleep appears central to our discussion, since cortical homeostasis presumably operates to maintain the efficiency of the higher brain centers. Hebb (1949) has suggested that loss of sensory input in sleep causes disruption of "phase sequences" (unified neural processes built up through learning, which result in organized thought) and some experimental studies of artificially induced sensory and/or perceptual deprivation appear to support the notion that operation of phase sequences depends upon a steady stream of sensory input. After some time has passed in the isolation chamber, with external stimulation sharply restricted, nearly all *S*s report a lack of clarity in thinking, with difficulty in directing their thoughts (Bexton, Heron, & Scott, 1954; Lilly, 1956).

Assuming that the quality of mentation may be indicative of the degree to which phase sequence, or organized thought, is operating at any given time, we might consider the nature of *mental content* reported on awakenings from both NREMS and REMS in an effort to shed further light on the relative effectiveness of cortical functioning in the two sleep phases.

As indicated, experimenters are in substantial agreement that *S*s recall mental content from *all* stages of sleep (Dement & Kleitman, 1957b; Goodenough et al., 1959; Orlinsky, 1962). Indeed, despite numerous changes found in brain functions during NREMS, in about 20% of awakenings from this sleep phase, *S*s report that they have been "awake" and "thinking" about matters of interest to them. Does this mean that these *S*s are effectively oriented toward reality at this time? Obviously this finding calls for careful evaluation since it is difficult to conceive of a neurophysiological mechanism by means of which ordinary secondary process thought could be maintained during a state of markedly lowered vigilance.

The occurrence of integrated thinking during NREMS seems questionable in light of recent evidence which suggests that NREM sleep may impede the consolidation of memory traces (Portnoff, Bakeland, Goodenough, Karacan, & Shapiro, 1966). Even for moderately organized thinking, short-term memory is essential, and it may therefore be significant that a vague, wandering, fragmented quality is frequently attributed to "thoughts" reported on awakenings from NREMS (Foulkes, 1962, 1964; Kamiya, 1961; Rechtschaffen, Verdone, & Wheaton, 1963).

While at present there are no entirely satisfactory answers to the question of whether or not "thinking" occurs during NREMS, the possibility should be considered that some, perhaps even the majority, of these thinking reports actually refer to mental operations occurring during the transitional period between sleep and waking. Support for such a hypothesis may be found in the experimental work of Goodenough et al., who have shown that thinking reports are more likely to occur following gradual awakenings from NREMS than following abrupt awakenings from this sleep phase (Goodenough, Lewis, Shapiro, Jaret, & Sleser, 1965).

While these data suggest that some NREM *thinking* reports may be reports of awakening (hypnopompic) experiences, NREM *dreams* occurred as often on abrupt as on gradual awakenings. It seems more likely, therefore, that these dream reports reflect some form of mental activity actually occurring during NREMS. Foulkes (1962) obtained dream reports in 54% of NREM awakenings. These dreams are, by definition, perceptual experiences, and we might wonder if they reflect some "need" for sensory stimulation or "input" during sleep. Such a possibility will be considered when we discuss the more full-blown, vivid dreams associated with the REM phase.

Data which might shed additional light on NREM thinking reports are the brief periodic lightenings of the EEG pattern, commonly noted throughout NREMS (Kleitman, 1963), suggesting fleeting arousals followed by rapid return to sleep. It seems plausible to assume that at these moments of EEG arousal some form of thinking may take place, with those thoughts remembered by *S* when he is awakened in the laboratory. If during the intervening period of NREMS the sleeper had been operating in a diffuse, disorganized mental stage, he might have no way of estimating just when these thoughts had occurred. It might therefore seem to him that he had been thinking just prior to being awakened.[3] Taking the above possibilities into account, we would interpret subjective reports of secondary process thinking during NREMS with caution at this point.

In between 26% (Foulkes, 1962) and 54% (Kamiya, 1962) of awakenings from NREMS, *S*s seem unable to recall any mental content whatever. This suggests the possibility that periods of total unconsciousness take place during NREMS; or, alternatively, that mentation of a nonverbal, nonperceptual nature occurs which cannot be recalled.

Some observations by Gastaut and Broughton (1965) on behavior following abrupt awakenings from NREMS shed further light on the level of cortical efficiency in this sleep phase. These investigators found that children awakened from Stages 3 and 4 sleep (usually only with great difficulty) showed extreme confusion, were amnesic with respect to the awakening experience, and displayed behavioral evidence of continued sleep even though they showed awake EEG patterns. The *S*s could answer questions only after intense stimulation and then in a weak voice and with simple abrupt answers. Such confusional reactions were never present during awakenings from initial Stage 1, from Stage 2, or from REMS; in fact, it is reported that *S*s awakened abruptly from REMS became immediately alert (Broughton, 1966). The confusional reactions reported in one-half of the normal children awakened from Stage 4 are consistent with the hypothesis that a breakdown of organized thought processes occurs in NREMS.

---

[3] Cf. James (1890) and Head (1923) for discussion of the tendency to preserve an illusion of mental continuity despite actual gaps in the stream of consciousness. Such illusions of continuity may take place following petit-mal seizures, for instance, or following fast-acting anesthetics.

Although mental content which is remembered on awakenings from NREMS appears to be "drifting," "disorganized," and "nondirected" (Rechtschaffen, Verdone, & Wheaton, 1963), mental content recalled from REMS is described as being vivid and *elaborately organized* (Foulkes, 1962; Kamiya, 1961; Rechtschaffen, Verdone, & Wheaton, 1963). Foulkes thus characterizes the transformation from NREM to REM sleep as progressing "from kaleidoscopic flux (NREMS) to continuity and integration (REMS)" (Foulkes, 1964) with a shift "from fragmentary and disjointed content to elaborated and integrated content" (Foulkes, 1962).

A novel approach to studying shifts in mentation between the two sleep phases is reported by Fiss, Klein, and Bokert (1966). Comparing TAT stories told on awakenings from REMS with those told on awakenings from NREMS, these investigators found that while blocking and relative paucity of ideation characterized the stories from NREMS (in 40% of such awakenings *S*s declaring, "I can't think") such blocking never occurred after any REMS awakening. By contrast, stories obtained after interrupted REM periods tended to be vivid, imaginative, bizarre, and emotional, with REM-awakened *S*s showing faster reaction times.

Along similar lines, Mandell, Mandell, and Jacobson (1965) report evidence of special mental alertness during REMS, in contrast to NREMS. When these investigators read digit series to sleeping *S*s, asking them to repeat the numbers upon being awakened 5 seconds later, *S*s were found to have superiority of recall for digits presented during REMS, while they failed to recall any of the digits presented during NREMS. In like manner, Cobb et al. have shown that *S*s can give accurate behavioral responses to complex meaningful suggestions administered during Stage-1-REM, but not to suggestions administered during NREMS (Cobb, Evans, Gustafson, O'Connell, Orne, & Shor, 1965).

Sleep Disturbances in NREMS

We have cited evidence which suggests that descent into NREMS is accompanied by progressive disorganization of the mental processes which ordinarily serve adaptation in the waking state. Since loss of vigilance presumably threatens any organism that depends on cortical alertness and perceptual orientation for survival, we might wonder whether anxiety or stress reactions are found in the deeper NREM sleep stages.

Homeostatic mechanisms are normally anticipatory in nature, forestalling pathologic conditions and thus largely obviating the necessity for stress reactions, but such responses are not always prevented. Perhaps different persons vary in the degree to which they can "risk" loss of awareness in NREM sleep, just as in waking life we find marked interindividual differences in tolerance to extreme sensory deprivation.

It may be noteworthy that physiological states suggestive of stress occur in certain *S*s concomitantly with lowering of vigilance in NREMS. Possible stress reactions noted at this time include episodes of sleepwalking which have been shown to commence exclusively in NREMS (usually in Stage 4) (Gastaut, Batini, Broughton, Fressy, Roger, & Tassinari, 1964; Jacobson, Kales, Paulson, Lehmann, & Zweizig, 1965a, 1965b) and which are frequently accompanied by signs of fear which according to Jacobson et al. (1965a) may vary from mild anxiety to outright terror on the part of the somnambulistic *S* (and during which *S*s frequently show increased or even doubled heart rate [Gastaut & Broughton, 1965]).[4] Also suggestive of stress are the sleep terrors of children observed to occur exclusively in Stage 4 sleep (Gastaut & Broughton, 1965); episodes of enuresis which have been shown to occur almost exclusively in NREMS, often in Stage 4 in the 15 minutes preceding REMS (Broughton & Gastaut, 1963; Hawkins, Scott, & Thrasher, 1965); and severe nocturnal anxiety attacks in adults (the classical "nightmares") which tend to take place either during NREMS or concurrent with an awakening from this sleep phase (Gastaut & Broughton, 1965).

Some of these reactions could also be interpreted as results of release from cortical inhibition during the deeper stages of sleep. Two tentative explanations are suggested for these disturbances: (*a*) Deafferentation in NREMS may lower thresholds for stress reactions in chronically anxious subjects; and (*b*) cortical depression in NREMS may lead to cortical release phenomena in certain predisposed *S*s. These possibilities are not mutually exclusive. Possibly there is an interaction between stress and release factors in the sleep disturbances.

In addition to pathological behavior in *S*s inclined towards episodic phenomena in sleep, even entirely normal *S*s seem especially vulnerable to bizarre reactions if awakened abruptly from NREMS. Thus Gastaut and Broughton have been able to provoke varying degrees and durations of automatic behavior (verbal, gestural, or even ambulatory, i.e., an induced sleepwalking attack); relative behavioral nonreactivity to exteroceptive stimuli; retrograde amnesia for intercurrent events and other "symptoms" of sleep disturbances in entirely normal persons, merely by the expedient of abrupt forced awakenings from Stage 3 or 4 sleep (Broughton, 1966). It may be pertinent to recall Dement's comment on NREMS, to the effect that "there is some small hint that in certain rare instances, the ultimate, inexpressible terrors may arise out of this time of possible vulnerability."[5] Dement cites one adult *S* who on two or three occasions abruptly terminated a period of NREM sleep in his laboratory by sitting bolt upright and emitting a "bloodcurdling scream" although there was no dream recall or even recall of screaming (Dement, 1964).

---

[4] Sleepwalking incidents lasting more than 20–40 seconds have been shown to be accompanied by a change in the EEG tracings to low voltage faster frequencies (Kales, Jacobsen, Paulson, Kales, & Walter, 1966). Our point of view suggests that such incidents could be partial "arousal" responses occurring in reaction to stress in NREMS. This however does not imply that NREMS is an activated state, but rather that it may be *interrupted*, under certain conditions, by episodes of partial activation.

[5] Dement referred specifically to the possible lowering of the "repression barrier" in deep NREMS.

THE REM PHASE

Cortical Activation during REMS

To summarize thus far: We have presented evidence suggesting that NREMS is characterized by functional deafferentation; lowered cortical tonus; reduced vigilance; general disorganization of associative processes (loss of "phase sequence") and, in certain *S*s, episodic phenomena which might be interpreted as indicating stress. We will now turn to REMS.

In contrast to NREMS, most areas of the neocortex are as highly active during REMS as they are in the *most alerted waking state* (Dement, 1964, 1965a; Roffwarg, Muzio, & Dement, 1966; Snyder, 1965b). During the rapid eye movement phase there is a fast, irregular, low voltage (desynchronized) EEG pattern resembling an awake pattern except for relative absence of alpha activity. At this time cortical blood flow increases (Kanzow, Krause, & Kühnel, 1962), and hippocampal theta rhythms (Shimazono, Horie, Yanagisawa, Hori, Chikazawa, & Shozuka, 1960), pyramidal tract activity, (Arduini, Berlucchi, & Strata, 1963), cortical direct current potentials (Wurtz, 1965), cortical induced movement thresholds (Hodes & Suzuki, 1965), and auditory evoked responses in animals (Herz, 1965), together with numerous other measures, show marked changes which bear a striking resemblance to the most attentive waking state.[6]

This neuronal storm is not accompanied by corresponding behavioral activation, however. Except for sporadic jerks or twitches, *S* remains limp during REMS. Nevertheless, cats with bilateral destruction of the *locus coeruleus* (a noradrenergic nucleus in the pons which appears to account for the muscular atonia of REMS) behave differently. During REMS these cats, while totally unresponsive to actual visual stimuli, are observed to stare fearfully at invisible objects, to stand on their feet and exhibit ragelike behavior, and even to "stalk" imaginary moving objects, in marked contrast to their quiet state during NREMS (Jouvet, 1965b). Thus, with specific motor-inhibitory centers removed, REMS gives indications of being an intensely activated state both behaviorally and neurophysiologically.

Another impressive characteristic of REMS is the massive activation in the sensory receiving areas of the brain which occurs in the relative absence of any external sensory input (Bizzi & Brooks, 1963; Brooks & Bizzi, 1963). Evarts (1962) reports that single neurons in the primary visual cortex of the cat typically have a 70% higher rate of discharge during REMS than during NREMS, and Bizzi (1965) has shown that the spontaneous discharge of large numbers of lateral geniculate neurons is modified during REMS, resulting in an increase in rate of firing in the majority of these cells. Even the modulations of tone in the fine muscles of the middle ear show phasic and tonic changes during REMS, just as they do in animals responding attentively to waking events (Baust, Berlucchi, & Moruzzi, 1964; Dewson, Dement, & Simmons, 1965).

Current data thus suggest that REMS may approximate a state of *endogenous afferentation,* and Roffwarg et al. (1966) have recently outlined a mechanism by which such afferentation may take place. These authors cite findings which suggest that bursts of impulses emanating from a "REM sleep center" in the pons ascend to the thalamus whence they course along the primary thalamo-cortical projection pathways (as environmentally-induced impulses do) to sensory receiving areas in the cortex; and Dement (1964) has suggested that once this "internal sensory input" is "substituted" in the "stimulus-response chain," higher centers interpret and react to it as if it were a set of true percepts impinging on the nervous system from without.

---

[6] While the above findings suggest that cortical tonus is enhanced during REMS, evidence is less clear-cut with respect to vigilance in this state. During REMS a rise in arousal thresholds is reported in cats (Jouvet & Jouvet, 1963) while lowered arousal thresholds are reported in adult humans (Rechtschaffen et al., 1966).

If REMS represents a state of endogenous sensory input, we might differentiate three states with respect to afferentation:

1. *Wakefulness* (predominantly exogenous afferentation) during which the organism regularly receives, interprets, and responds to external stimuli.

2. *NREMS* (functional deafferentation) during which the organism may receive external stimuli, but will interpret and respond to them only in rudimentary fashion, if at all.

3. *REMS* (predominantly endogenous afferentation) during which the organism is largely unresponsive to external stimuli, while at the same time a dramatic increase in internal stimulation occurs, with the presence of vivid dreaming in humans suggesting that this input is patterned and meaningful.

In view of the polarity of the two sleep phases with respect to afferentation, we have considered the possibility that REMS may have the functional effect of counteracting deafferentation in NREMS. More broadly, it might be seen as reinvigorating the cortex following depletion of cortical functions in NREMS.

The dreaming process, an integral part of human REMS and one perhaps common to many lower animals as well, (Dement, 1965a; Snyder, 1965a) may be an important component of this endogenous afferentation. In order to examine this possibility we will consider evidence suggesting that REM dreaming, as an hallucinatory process, provides the sleeping brain with *patterned sensory input.*

Hallucinations as Replacement Phenomena

Hallucinations (usually either visual or auditory) have been reported to occur in certain persons under conditions of experimental sensory deprivation (Freedman, Grunebaum, & Greenblatt 1961; Vernon, McGill, Gulick, & Candland, 1961) and apparently it is not unusual for persons subjected to reduced sensory input to greet these visions as a form of welcome relief from monotony. In similar manner, explorers, prisoners, and others in states of natural isolation are frequently reported to welcome their hallucinations (Miller, 1962). It would seem that with diminished stimulation, stimulation of *any kind* may be avidly sought. Lilly (1956) for example, describes a "stimulation-action lust" so intense in *S*s undergoing sensory deprivation that stroking one finger with another becomes an ineffable pleasure.

We might wonder whether those hallucinatory experiences which are greeted with relief or pleasure under conditions of reduced sensory input have this effect because they perform a necessary *replacement function.* If so, an organism deprived of sensory input might be regarded as creating its own input, to which it subsequently attends.

Evidence relating to hallucinatory

experiences in patients deprived of sensory input through illness or injury tends to support the notion that replacement of absent sensory input may be a function of hallucinations. Bartlett (1951), in his description of an old man with cataract, showed that the hallucinatory activity occurred only in that limited portion of the visual field which corresponded to the position of the cataract; while observations of phantom-limb experiences are equally specific in showing strictly limited hallucinosis suggestive of replacement phenomena (Bartlett, 1951; Fisher, 1962). Similarly, auditory hallucinations may occur in individuals having a progressive loss of hearing, with musical perceptions frequently noted in such patients (Rosanski & Rosen, 1952).

Still further, schizophrenia, a condition often accompanied by hallucinations, has recently been compared to a state of *sensory deprivation* (Rosensweig, 1959; Searles, 1963). As Searles indicates, schizophrenic patients may become so isolated from their environment that they scarcely perceive persons or objects in the natural world. A view of schizophrenia as a form of sensory deprivation suggests that the hallucinations of this illness might be seen, at least in part, as results of the sensory-deprived state of the patient. If so, the hallucinations might represent a pathologic attempt to counteract the deprivation through creating an illusory world of the senses.

In similar manner, we might view the experience of *dreaming* as a form of recovery following functional deafferentation in NREM sleep. Just as the REM state is characterized neurophysiologically by activation of the sensory receiving areas of the brain; on the psychological level, the dream process (at least in man) appears to represent a form of intricately *patterned perception*. An elaborate structuring of the spatio-temporal field takes place during dreaming, and there is evidence that the length of the dream is roughly proportional to the hallucinated dream events occurring within it (Dement & Wolpert, 1958). Furthermore, subjective evidence indicates that early sequences which form part of a long dream are often recalled by the dreamer during a later sequence of the same dream; just as in waking life we typically remember events leading up to the present. Such references to the "past" of the dream, experienced *during* the dream, suggest that a complex memory process operates during REMS, and, as we have seen, there is also some objective evidence of ability to commit events to memory at this time (Mandell et al., 1965). Similarly, juxtaposition of dream objects in space, and perception of motion in the dream, suggest a sophisticated (though still somewhat distorted) orientation in space during REMS. Evidence thus points to an extensive recovery of integrative functions during the rapid eye movement dream, seemingly in contrast to the disintegrative trends noted in NREMS [7] (see above).

IMPLICATIONS OF A HOMEOSTATIC HYPOTHESIS

The Sequence NREMS/REMS

If, as suggested, REMS is triggered in reaction to loss of cortical tonus in NREMS, then it should ordinarily *follow* NREMS. This appears to be the case. Normal mature human *S*s, for example, do not show REMS until roughly 60 to 90 minutes have passed after sleep onset, and an initial Stage-4–Stage-1–REM contingent relationship has been postulated by Webb and Agnew (1965, 1966) based on these investigators' normative studies of sleep patterns in young adults. Similarly, a period of NREMS has been shown to precede REMS in mature animals[8] (Levitt, 1965; Roldan, Weiss, & Fifkova, 1963; Sawyer, 1965) except where the sleep cycle has been radically altered by manipulation as in REMS-deprivation.

---

[7] It would seem that NREM dreams may fulfill a sporadic afferenting function, similar to that of the REM dream, although they are milder and not accompanied by the physiological changes of the REM state. Perhaps the intermittent sensory "input" which they afford contributes toward the maintenance of a minimal level of cortical tonus during NREMS which is not sufficiently pronounced to register by electrophysiological means.

NREMS also appears to represent an indispensable condition for the occurrence of artificially stimulated REMS phenomena. Jouvet and Michel (1960a, 1960b) have demonstrated, for example, that in the intact cat it is possible to induce REMS by stimulating certain points in the midbrain or pons *provided this is done during NREMS*. Furthermore, having once stimulated REMS in this manner, it was impossible for these investigators to repeat the effect for 15–30 minutes. This refractory interval suggested to them that REMS is due to the liberation of some neurohumoral agent for whose production and accumulation NREMS, or something associated with it, provides the stimulus.

Sleep experimentally produced by Hernández-Peón (1964) through chemical means, and by Sawyer (1965) through electrical stimulation, also has not led directly to REMS, NREMS always occurring first and giving way to a subsequent REM phase. Also of interest, a study by Sovrad and Kamarova (1965) has shown that rats injected with bulbocapnine (which supresses the amplitude of the high-voltage slow waves of NREMS) show no episodes of REMS until the amplitude of the high-voltage sleep waves has returned to base-line level (i.e., until sufficient "depth" or "intensity" of NREM sleep has been restored). NREMS thus appears to be a prerequisite for REMS unless certain types of unusual experimental manipulation or pathological states interfere with the normal cycle.

Narcoleptics tend to commence sleep with REMS, and Rechtschaffen, Wolpert, Dement, Mitchell, and Fisher (1963), have suggested that this illness represents a specific disturbance of the pontine system leading to pathological over-activity or premature triggering of the REM sleep center in the pons. A prominent EEG feature in narcolepsy is the frequent appearance of drowsiness and sleep patterns during the day, even when *S* appears to be wide awake and alert (Hishikawa & Kaneko, 1965); and Hess (1965) points out that "the narcoleptic's brain may be asleep while he is walking, working, etc." Possibly, therefore, we might be justified in viewing the narcoleptic as experiencing an abnormal degree of functional deafferentation during waking life. While the explanation for sleep onset REMS in narcoleptics appears to lie in the realm of neurophysiological pathology, conceivably a contributing factor might be an urgent need for endogenous afferentation created by the persistent symptoms of drowsiness seen in these patients. According to the hypothesis here presented, a mounting need for afferentation during these states of lowered vigilance might contribute to the likelihood that rapid eye movement sleep would be experienced prematurely.

Neonates also typically enter the REM state at the beginning of sleep (Roffwarg, Dement, & Fisher, 1964; Roffwarg et al., 1966), and according to Roffwarg[9] the occurrence of sleep-onset REMS in infants seems to reflect a generally greater "need" for REM sleep in the infant organism which this author has related to the requirements of the maturing brain for endogenous sensory stimulation, since at that period there is relatively little opportunity for exogenous input.

---

[8] Weitzman, Kripke, Pollak, & Dominguez (1965) report occasional sleep onset REMS in "preadolescent" monkeys, however.

Some findings by Jouvet and his group (Jouvet, 1965a) also have bearing on the NREMS/REMS sequence. These investigators have demonstrated that decorticate (pontile) cats regularly show periods of muscular atonia together with phasic activity (including rapid eye movements) similar to that seen during REMS in intact cats, while these pontile animals show no NREMS. Despite the absence of cortex, the REMS system appears to remain in operation as long as its pontine triggering mechanism remains intact. While these investigators seem clearly to have demonstrated an independence of REMS from NREMS in their decorticate preparations, we need not assume that such an independence obtains in unoperated animals. As previously suggested, it may be that where there *is* a cortex, the occurrence of REMS serves to restore cortical functions following lowered cerebral vigilance in NREMS. It may be pertinent in this connection to recall that animals that naturally have little neocortical tissue, such as certain reptiles, do not show any clearly identifiable REM sleep.

An Accumulable Tendency toward REMS

Influences which tend to deepen sleep, such as some drugs (Feinberg, Koresko, Gottlieb, & Wender, 1964; Hendley, 1963) and sleep deprivation (Berger & Oswald, 1962), may have the immediate effect of *decreasing* the amount of REMS (although typically REMS is later elevated above baseline levels in apparent compensation for the earlier suppression). Thus REMS is sometimes lessened when cortical tonus reaches exceptionally low levels, suggesting temporary postponement of the hypothesized homeostatic regulation. This may occur because under certain conditions (such as excessive fatigue) NREMS takes precedence over REMS, antagonizing the physiological mechanisms responsible for it (Roffwarg et al., 1964). Since all homeostatic mechanisms can be thwarted by some circumstances, perhaps it might be pertinent to distinguish between the effects of normal deep sleep on the one hand, and the possible effects of deep sleep due to an excess of noxious factors (extreme fatigue, drugs, etc.) on the other, since it is the latter conditions which appear capable of upsetting the normal REMS cycle.

Evidence suggests, however, that postponement of the hypothesized homeostatic responses may also occur under conditions which can be considered commonplace, as, for example, when a waking state intervenes. Webb[10] reports on *S*s brought into the laboratory 2 hours after a full night's sleep (9 A.M.) and permitted to sleep for 2 hours. These *S*s displayed remarkably elevated amounts of REM sleep during these morning naps (around 40%), similar to the concentration of REMS seen during the morning hours of uninterrupted night sleep, while they showed almost no Stage 3-4 sleep, often bursting into REMS immediately.

Triggering of REMS by directly preceding NREMS could not have occurred in this case. However, we might explain these results in terms of the homeostatic hypothesis by assuming that a prolonged process of reafferentation in REMS was temporarily interrupted by the waking interval. While the waking interval itself might be viewed as continuing to reafferent the cortex, evidence suggests that a good deal of waking stimulation is necessary before cerebral efficiency is fully restored after sleep.

---

[9] H. P. Roffwarg, personal communication, July 27, 1965.

[10] W. B. Webb, personal communication, June, 1966.

In general, a person's performance seems to be worse after getting up in the morning than before going to bed the night before. When Omwake (1932) submitted *S*s to simple sensory-motor and intellectual tasks before they retired for the night, and again after they had had 2, 4, 6, or 8 hours of sleep, respectively, she found a marked deterioration on all tests after 2 hours of sleep as compared to the scores obtained before going to bed. There followed a gradual improvement as the period of sleep was lengthened, but even after 8 hours of sleep, the performance was not up to the presleep level.[11] It is conjectured, therefore, that the NREM deafferenting process may either overshoot its mark during sleep, or conversely, that reafferentation through REMS does not fully counteract loss of cortical tonus by the end of the night.

If residual effects from the night's deafferentation which persist into the postsleep morning hours are postulated, reafferentation may be an important physiological requirement following awakening in the morning, with an individual who returns to sleep before this hypothetical deficit is met encountering a cortical demand for intensive sensory stimulation in REMS.

Temporary postponement of the REM response as described above clearly implies an accumulable tendency to produce REMS. Because of its importance for an understanding of the proposed homeostatic principle, such a possibility will now be considered.

According to the "homeostatic" theory, the neocortex regularly calls upon centers in the pons to supply it with necessary stimulation during sleep. We might conceive of a functionally deafferented cortex sending messages to a REM sleep center in the pons demanding endogenous afferentation, which is then effected by means of impulses sent coursing along ponto-thalamo-cortical circuits to the sensory receiving areas of the brain. Such a process could utilize the equivalent of intracortical sense organs (as Davis, 1950, has suggested) capable of responding to the level of cortical tonus even during deepest NREMS; or it might operate in some other as yet unsuspected fashion to monitor cerebral excitation.

It seems reasonable to suppose, however, that a homeostatic mechanism of such presumed importance for the proper functioning of the brain would not rely upon immediate feedback alone, but have its continuing operation insured through some biological timing mechanism essentially resistant to change. Jouvet and Michel (1960a, 1960b) and Dement (1965b) have offered evidence which suggested the accumulation of some biochemical substance (probably of a neurohumoral nature) which triggers REMS, and it would seem that the notion of a biochemical triggering mechanism may have value for the proposed homeostatic hypothesis. A biochemical triggering substance is by no means the only explanation for an accumulable tendency to have REMS, however. As Hartmann (1965) indicates, a combination of substances, the lack of a substance, or altered sensitivities of certain cell groups might explain the facts as well.

---

[11] It is interesting that Omwake's subjects showed the greatest loss of cerebral efficiency after the first 2 hours of sleep where NREMS usually predominates, with a progressive improvement in task performance noted over the portions of the night where "reafferentation" through REMS might be most likely to occur.

Whatever the exact nature of such a process, eventual restoration of cortical tonus in the face of temporary postponement implies a mechanism capable of registering the effects of loss of tonus and retaining this information over a period of time. One possibility that might be considered is that lowering of vigilance and consequent deafferentation in NREMS automatically permits accumulation of a REMS-triggering substance, with eventual onset of REMS ensured once this substance reaches a preset level of concentration. While a REMS "potential" would thus be built up during deafferented states, onset of REMS could be deferred under conditions not favoring immediate endogenous afferentation. Such might be the case, for example, at the beginning of the night when the need for slow-wave sleep may preempt REMS, when a waking state intervenes, or when REMS is artificially prevented by experimental manipulation.

A mechanism whereby lowered vigilance causes accumulation of a REMS triggering substance could have the advantage of sensitively balancing the degree of eventual cortical recovery with the degree of previous loss of cortical tonus, while at the same time permitting delay in timing of the homeostatic response. Conceivably such a process might have been operating in Webb's napping *S*s. A concentration of deafferenting sleep immediately prior to the onset of REMS would not have been required to trigger the REMS process according to such a hypothesis, since the triggering substance could have accumulated during the previous night's sleep.

Cortical Homeostasis in Human Sleep

Since most homeostatic influences depend upon numerous values that delicately balance each other, it becomes an awesome task to attempt to gauge such a process. Any theoretical model must fail to do justice to the interplay of factors to which the cortical homeostatic process presumably responds. Since not all these factors can be handled at any one time, certain regulatory tendencies will presumably be superseded by other tendencies which bring about more urgent adjustments. A deficit state requiring homeostatic regulation should generally persist, however, even though the corrective reaction to it is temporarily thwarted.

Evidence suggests that a simple action-reaction model does not describe the manner in which cerebral homeostasis operates during a night of prolonged sleep such as is seen in the adult human. Instead, what we may be seeing is a competition between conflicting homeostatic responses. While certain correlations are reported between the *length* of NREM periods and the length of their subsequent REM periods in man (Hartmann, 1965; Rechtschaffen & Verdone, 1964), there are no similar indications of any one-to-one relationship between the *degree* of loss of cortical tonus in a NREM phase and either the degree of cortical activation evidenced in its immediately following REM phase or the duration of the latter. For example, the initial NREM phase which follows sleep onset in man might be described as his "deepest" sleep of the night because of the preponderance of Stage 4 at this time, yet this sleep phase is followed by the shortest and often the physiologically least activated of the night's REM periods. To understand this occurrence, it may be useful to view the alternation of the sleep phases in the broader context of the sleep-waking cycle.

As Davis (1950) and Gellhorn (1953) indicate, cortical homeostasis presumably operates to reduce cerebral overexcitation as well as underexcitation. If we were to consider the awake state as leading toward eventual overstimulation, and the sleeping state as leading toward eventual understimulation, we might speculate that not only must cortical efficiency be maintained during functionally deafferented states by a process of endogenous afferentation, but conversely, that following states of *over*afferentation functional deafferentation becomes necessary.

Conceivably, after prolonged sensory bombardment, occlusion of sensory input is required to reduce cortical stimulation to levels more acceptable to the organism. If so, a homeostatic reaction to overafferentation might be a significant factor in the initial phases of normal sleep.

If we view the sleep of the first NREM phase as serving to counteract the effects of prolonged wakefulness, we might not expect the first REM period to afford extensive endogenous afferentation (as evidenced by its level of activation or duration). It would seem rather that following accumulated intensive sensory bombardment from the day, overafferented neural structures might resist further stimulation (by REMS *or* by waking) until such time as sufficient recovery of such structures has taken place. Possibly, therefore, the more vigorous manifestations of REMS might be called into play only after the effects of waking stimulation have been fully counteracted. This would seem likely to be the case during the latter portions of the night. Man's diurnal pattern of uninterrupted daytime wakefulness might therefore be thought of as placing a special burden on cortical homeostatic regulation.

This possibility appears to receive some support from the fact that the distinct difference between the early and late portions of the night is first seen at that period in the child's development when daytime napping has finally terminated and an overall diurnal rhythm of sleep and wakefulness becomes established. (Roffwarg, Dement, & Fisher, 1964). Also possibly relevant, Maron, Rechtschaffen, and Wolpert (1964) report that nocturnal sleepers who napped in the afternoon had shorter REM latencies than sleepers who napped in the evening. Roffwarg, Muzio, and Dement (1966) interpret these various findings by suggesting that "fatigue" may develop under circumstances of prolonged arousal which tips the normal balance between REMS and NREMS, augmenting temporarily the "need" for deep NREM sleep and simultaneously antagonizing REM sleep processes.

One might further conceptualize that man's advanced wakefulness capacity of 15–17 hours results in mild, habitual deprivation of sleep which might otherwise occur spontaneously during the day. Such intermittent sleep is frequent in infants and lower animals that are not trained to resist the sleep impulse. In these organisms, functional deafferentation may occur more or less whenever it is required. If so, they should not approach sleep with the same "need" for functional deafferentation as adult man and we do find them experiencing relatively short NREM–REM cycles.

A brief, frequently unstable initial REM period does occur in man, however, during his first hours of sleep. We conjecture that the pacing of REMS depends upon a phylogenetically old and biologically primitive mechanism which regularly triggers this state in response to lowering of cerebral vigilance in NREMS, perhaps thereby serving the adaptive function of periodically alerting the organism during sleep for "sentinel" purposes as suggested by Snyder (1966). Mechanisms initiating activated sleep may thus be automatically called into play by the first NREM phase even though the central nervous system may not yet be sufficiently unburdened at this time to tolerate extensive endogenous afferentation. This would be likely to result in a relatively short first REM period.

A process similar to that outlined above might be expected to lead to a complex sequence of actions and reactions between sleep phases. While not necessarily demonstrating precise compensatory relationships between adjacent phases, such a homeostatic process may nevertheless result in gradual approximation, over the night, to a "center-of-gravity" position with respect to the degree of cortical tonus maintained during sleep and subsequently the degree of cerebral efficiency available to the sleeper upon awakening in the morning. A possibility suggested by Omwake's (1932) experiment is that this equilibrium may not, in fact, be fully achieved until several hours of morning activity have supplied further stimulation.

The model described above would seem in accord with the homeostatic process as it operates in other areas and as it is paralleled in servo-mechanisms, where we may see the counteractivity which corrects the undesirable deviation proceeding in an oscillating, irregular fashion, sometimes overdoing and sometimes underdoing the job it is called upon to perform, until some measure of equilibrium is achieved.

Functional Deafferentation in Stage 2

We have postulated two processes or tendencies during sleep, the one leading toward organismic rest and deafferentation, and the other toward activation and afferentation. While these tendencies have seemed linked to certain sleep stages, this linkage may not be absolute. It is conceivable, for example, that a general tendency toward afferentation (or deafferentation) might persist despite temporary elimination of the specific EEG sleep stage with which it is normally associated. Some evidence which bears on this question is reported by Fisher, Gross, and Byrne (1966) and Karacan and Goodenough (1966) who indicate the persistence in NREMS of a cycle of penile erection in human males despite REMS deprivation. If the REM-period erection were to be viewed as a form of organismic activation indicative of the energizing of a major sensory receptor area, these findings suggest that with REMS eliminated, NREMS may assume some of the enlivening or "afferenting" functions ordinarily linked with the REM state.

Reasoning along the same lines, we might wonder if, under certain circumstances, Stage 2 could assume functions such as profound deafferentation, which are usually associated more directly with Stages 3 and 4. As previously indicated, *S*s in slow-wave sleep (Stages 3–4) tend to respond only minimally to exteroceptive stimuli with this functional deafferentation maximalized in Stage 4. While it might be expected, therefore, that these deep NREM sleep stages would be likely to bring about an organismic "need" for REMS, evidence suggests that Stage 2, combined with only minimal amounts of Stage 3, may also perform such a function.

In a recent experiment Webb (see

Footnote 10) deprived *S*s for 7 nights of most of Stage 3 and all of Stage 4 sleep by giving a warning tone whenever slow waves appeared on the EEG, followed by an electric shock to the foot if Stage 3 continued. The resulting records showed a greatly increased Stage 2 percentage with REM percentage not reduced significantly below base-line levels. These results might be interpreted as indicating that Stage 2 sleep alone is sufficiently deafferenting to require endogenous afferentation through REMS, an explanation which seems plausible in light of experimentation showing functional deafferentation in the visual system at sleep onset (Rechtschaffen & Foulkes, 1965). On the other hand, we cannot be certain that the prolonged Stage 2 sleep seen during Stage 3–4 deprivation is the same, qualitatively, as Stage 2 sleep experienced under normal conditions. If we assume that the shocks administered to *S*s do not eliminate the physiological requirements for Stages 3 and 4 sleep, but only the manifestations of these sleep stages (as evidenced on the EEG), we might wonder whether the increased Stage 2 seen in this experiment may not have assumed some of the functions of the eliminated sleep stages.

That the NREM sleep obtained was behaviorally "deep" despite the predominance of a Stage 2 EEG is suggested by the fact that *S*s tolerated a vast number of electric shocks during this phase (sometimes as many as 300–400 shocks per night, occasionally in intensities which would be tetanizing) yet upon awakening they did not report any subjective impression of having been shocked during sleep. This could, of course, be due to lack of consolidation of memory traces in NREMS, but the fact that *S*s did not readily awaken with these surprisingly heavily concentrations of shock appears to suggest that a considerable degree of deafferentation was effected during their NREM sleep despite the enforced lightening of the EEG.

REMS and Convulsive Phenomena

To turn to consideration of the mechanisms by which REMS may effect "toning up" of the CNS; Gellhorn (1953) suggests that convulsions may serve to restore cortical functions under certain circumstances. Perhaps it is significant, therefore, that REMS is characterized by paroxysmal phenomena (twitchings, bursts of eye movements, phasic CNS activity) which in certain respects resemble seizure activity. These convulsive aspects of REMS are particularly noticeable when REM sleep is first permitted following prolonged REMS deprivation. At such times animals may show twitching of the body, paws, and whiskers so intense that they occasionally present a picture of epileptic seizures (Jouvet, 1965a).

Reasoning from neurophysiological evidence, Evarts (1964) appears to adopt the position that the neuronal hyperactivity which characterizes REMS shows certain similarities to the paroxysmal discharge of epilepsy. Also examining the relationship of REMS to seizures, Zarcone, Gulevich, and Dement (1966) report that psychiatric patients, after a series of electroconvulsive shocks, showed marked attenuation of REMS for a number of days with no tendency to make up for that sleep phase later on. In a related study (Cohen & Dement, 1966) cats deprived of REMS, but receiving daily convulsions, showed a reduction in REMS-time to 50% of base-line levels, while they did not show any recovery rise such as that seen in their "deprived-only" controls. The investigators suggest that in these instances the convulsive experience may either have substituted in some fashion for rapid eye movement sleep, or else that the convulsions caused temporary dysfunction of that which mediates REMS.

Still other findings hint at a relationship between REMS and convulsive phenomena. Testing seizure thresholds, Cohen and Dement (1965) report a marked drop in electroconvulsive threshold in rats after a 6-day period of REMS deprivation; and Winters and Spooner (1965), noting that very short REMS tend to occur just prior to and after induced seizure bursts in narcotized cats, have raised the question of a possible relationship between the REM state and generalized nonconvulsant epileptic states.

In light of the foregoing, it is conceivable that the mechanism by which hypothesized cortical recovery in REMS takes place may in some way be related to that which brings about cortical recoveries following certain seizures. If so, the relationship between cortical homeostasis and the occurrence of seizures, proposed by Gellhorn, might also be applied to the REMS process. Reasoning in this manner, electroconvulsive shock might be seen as supplying such intense stimulation to the cortex that a prolonged refractory period ensues during which intensive stimulation by REMS is less possible, and perhaps also less required from a neurophysiological standpoint.

Related Evidence

What evidence is there to suggest that the extent of the hypothetical recovery in REMS reflects the extent of deficit brought about by NREMS?

Two experiments may be pertinent, although neither was designed with this question in mind. Hartmann (1965), studying the relationship between sleep phases, found a positive correlation between the length of NREM periods (in humans) and the length of their subsequent REM periods, while no correlation was found between the lengths of REM periods and their subsequent NREM periods.[12] He thus concluded that a need for REMS builds up during NREMS so that, other things being equal, a long NREM period is followed by a longer REM period. Along similar lines, a study by Rechtschaffen and Verdone (1964) shows a positive correlation between the amount of NREMS time prior to the onset of the first REM period of the night and the length of the first REM period.

One possible approach to the study of the interrelationship of the sleep phases might be to vary the amount of exogenous stimulation introduced into the sleep cycle during NREMS, and observe the effect of REMS. If greater amounts of exogenous stimulation are made available (in brief awakenings) and cortical tonus thus maintained at an exceptionally high level during NREMS, we might expect that there would be less need for hypothetical recovery of cortical tonus through REMS. While this has not been systematically studied, an experiment by Rechtschaffen and Verdone (1964) provides some data which may be relevant. These investigators showed that the typical "first-night effect" in the sleep laboratory (when *S*s show decrements in REMS-time as compared with subsequent nights) is related to the greater number of brief awakenings which also occur on the first night. Since such multiple awakenings seem to represent a generally heightened level of vigilance (probably due to the unfamiliar conditions of the experiment), one interpretation of these results might be that a need for restoration of cortical functions through REMS was rendered less necessary by the already alerted state of *S*s.

---

[12] Weitzman et al. (1965) report contrasting findings in monkeys, however.

Also relevant to the present hypothesis, a study by Roldan, Weiss, and Fifkova (1963) suggests a recurrent homeostatic interplay between NREMS and REMS. Studying the rat, these authors identified a well-defined cyclic sequence of EEG patterns repeated regularly throughout sleep. This sequence consisted of three consecutive stages: NREMS, REMS, and a brief arousal reaction regularly occurring following REMS. By systematic regularly repeated electrical stimulation of the mesencephalic reticular formation, or acoustic stimulation, it was possible to demonstrate that in the course of spontaneous sleep, cyclic changes in excitability paralleled the cyclic EEG changes. A gradual decrease in excitability in the reticular activating system occurred during the NREM phase, dropping out altogether in the pre-REM period and during REMS. Following the brief arousal period on termination of REMS, the entire cycle recommenced with greatly increased excitability displayed during the first stages of the new NREM phase. In keeping with the findings of Roldan et al., Snyder[13] reports a similar cycle of changes in pulse and respiratory rates in the opossum which regularly reach peak levels in the brief arousal period following REMS, gradually falling once more in the NREM period.

One interpretation of the findings of Roldan and his associates might be that REMS serves to restore cortical functions in these animals to a degree sufficient to produce the brief arousal and subsequent lowered arousal thresholds of the post-REMS period.

To turn now to the relationship between REMS and deafferented states, a means of studying this problem might be to investigate the effects of experimentally reduced sensory input on REMS. While there has been no research to date on the effects on REMS of sensory and/or perceptual deprivation, Wood has explored the effects of *social isolation* on sleep patterns (Wood, 1962). This investigator locked *S*s in a room empty save for two chairs, a table, and reading matter, permitting them no contact with other persons for the duration of their waking hours. The percentage of time spent in REMS on the night following isolation was then measured in the sleep laboratory. Results showed that REMS time on the postisolation night was approximately 60% greater than the mean percentage REMS time for four base-line nights ($p = <.01$), in all *S*s studied; and subsequent studies have shown this result to be unrelated to compensation for "first-night effect."[14] An increase in REMS following social isolation (which in this experiment included incidental partial sensory deprivation) is consistent with the hypothesis that REMS serves to supply endogenous sensory input in the absence of adequate input from the environment.

Implications for Research

In closing, we would suggest that experimenters working on the psychophysiology of sleep devote attention to the recurrent interplay between the sleep phases. To date, relatively little work has been done on this aspect of sleep research, with of course certain exceptions (Agnew, Webb, & Williams, 1964; Roldan et al., 1963; and a few others), most investigators focussing their attention on studying the sleep phases as discrete phenomena.

It would seem that research attempting to relate the phases to each other may be fruitful. For example, attention to the alternate lowering and raising of cortical tonus in NREMS and REMS, respectively could lead to a comparison of these sleep states on a number of variables related to cortical efficiency. Such investigations should bring about a more profound understanding of these sleep forms than the study of either in isolation, particularly if attention is directed to the continuum represented by the NREMS-REMS cycle (as in the study by Roldan et al., 1963). In addition, attempts to measure broader tendencies toward afferentation or deafferentation as these may fluctuate throughout sleep (frequently without regard for the boundaries of the various EEG sleep stages) might lead to useful predictions concerning certain phenomena, such as reporting of dreams from NREMS, which at present seem "out of phase" and difficult to assess with respect to their role in sleep.

---

[13] Frederick Snyder, personal communication, June 10, 1965.

[14] P. B. Wood, personal communication, March, 1966.

Finally, exploring the effects of prolonged, radically reduced sensory input on the incidence of REMS may help to clarify the relationship of rapid eye movement sleep to cerebral requirements for sensory stimulation.

REFERENCES

Agnew, H. W., Webb, W. B., & Williams, R. L. The effects of stage four sleep deprivation. *Electroencephalography and Clinical Neurophysiology*, 1964, **17**, 68–70.

Arduini, A., Berlucchi, G., & Strata, P. Pyramidal activity during sleep and wakefulness. *Archives Italiennes de Biologie*, 1963, **101**, 530–544.

Aserinsky, E., & Kleitman, N. Regularly occurring periods of eye motility and concomitant phenomena during sleep. *Science*, 1953, **118**, 273–274.

Aserinsky, E., & Kleitman, N. A motility cycle in sleeping infants as manifested by ocular and gross bodily activity. *Journal of Applied Physiology*, 1955, **8**, 11–18.

Barcroft, J. *The brain and its environment.* New Haven: Yale University Press, 1938.

Bartlett, J. E. A. A case of organized visual hallucinations in an old man with cataract and their relation to the phenomena of the phantom limb. *Brain*, 1951, **74**, 363–373.

Baust, W., Berlucchi, G., & Moruzzi, G. Changes in the auditory input during arousal in cats with tenotemized middle ear muscles. *Archives Italiennes de Biologie*, 1964, **102**, 675–685.

Berger, R. J. Tonus of extrinsic laryngeal muscles during sleep and dreaming. *Science*, 1961, **134**, 840.

Berger, R. J., & Oswald, I. Effects of sleep deprivation on behavior, subsequent sleep and dreaming. *Journal of Mental Science*, 1962, **108**, 457–465.

Bexton, W. H., Heron, W., & Scott, T. H. Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology*, 1954, **8**, 70–76.

Bizzi, E. Discharge patterns of lateral geniculate neurons during paradoxical sleep. Paper read at Association for the Psychophysiological Study of Sleep, Washington, March 1965.

Bizzi, E., & Brooks, D. C. Functional connections between pontine reticular formation and lateral geniculate nucleus during deep sleep. *Archives Italiennes de Biologie*, 1963, **101**, 666–681.

Bremer, F. Cerveau "isolé" et physiologie du sommeil. *Comptes Rendus des Séance de la Société de la Biologie et de ses Filiales*, 1935, **118**, 1235–1241.

Bremer, F. Effets de la déafférentation complète d'une region de l'écorce cérébrale sur son activité électrique spontanée. *Comptes Rendus des Séance de la Société de la Biologie et de ses Filiales*, 1938, **127**, 355–359.

Brooks, D. C., & Bizzi, E. Brain stem electrical activity during deep sleep. *Archives Italiennes de Biologie*, 1963, **101**, 648–665.

Broughton, R. J. Observations on the pathophysiology of certain paroxysmal sleep disorders. Paper read at Association for the Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

Broughton, R., & Gastaut, H. Étude polygraphique de l'énurésie nocturne. *Revue Neurologique*, 1963, **109**, 246–247.

Cannon, W. B. Organization for physiological homeostasis. *Physiological Review*, 1929, **9**, 399–431.

Cobb, J. C., Evans, F. J., Gustafson, L. A., O'Connell, D. N., Orne, M. T., & Shor, R. E. Specific motor response during sleep to sleep-administered meaningful suggestion: An explanatory investigation. *Perceptual and Motor Skills,* 1965, **20**, 629-636.

COHEN, H. B., & DEMENT, W. C. Sleep: Changes in threshold to electroconvulsive shock in rats after deprivation of "paradoxical phase." *Science,* 1965, **150**, 1318–1319.

COHEN, H. B., & DEMENT, W. C. Suppression of rapid eye movement sleep in the cat following electroconvulsive shock. Paper read at Association for the Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

COLEMAN, P. D., GRAY, F. E., & WATANABE, K. EEG amplitude and reaction time during sleep. *Journal of Applied Physiology,* 1959, **14**, 397–400.

DAVIS, H. Homeostasis of cerebral excitability. *Electroencephalography and Clinical Neurophysiology,* 1950, **2**, 243–247.

DEMENT, W. C. The effect of dream deprivation. *Science,* 1960, **131**, 1705–1707.

DEMENT, W. C. Experimental dream studies. In J. Masserman (Ed.), *Science and Psychoanalysis.* Vol. 7. New York: Grune & Stratton, 1964. Pp. 129–162.

DEMENT, W. C. An essay on dreams. In *New directions in psychology* II. New York: Holt, Rinehart & Winston, 1965. Pp. 135–221. (a)

DEMENT, W. C. Recent studies on the biological role of rapid eye movement sleep. *American Journal of Psychiatry,* 1965, **122**, 404–408. (b)

DEMENT, W. C., & KLEITMAN, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology,* 1957, **9**, 673–690. (a)

DEMENT, W. C., & KLEITMAN, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *Journal of Experimental Psychology,* 1957, **53**, 339–346. (b)

DEMENT, W. C., & WOLPERT, E. A. The relation of eye movements, body motility, and external stimuli to dream content. *Journal of Experimental Psychology,* 1958, **55**, 543–553.

DEWSON, J., DEMENT, W. C., & SIMMONS, F. Middle ear muscle activity in cats during sleep. *Experimental Neurology,* 1965, **12**, 1–8.

DUSTMAN, R. E., & BECK, E. C. Phase of alpha brain waves, reaction time, and visually evoked potentials. *Electroencephalography and Clinical Neurophysiology,* 1965, **18**, 433–440.

EVARTS, E. V. Activity of neurons in visual cortex of cat during sleep with low voltage fast EEG activity. *Journal of Neurophysiology,* 1962, **25**, 812–816.

EVARTS, E. V. Temporal patterns of discharge of pyramidal tract neurons during sleep and waking in the monkey. *Journal of Neurophysiology,* 1964, **27**, 152–171.

FAVALE, E., LOEB, C., MANFREDI, M., & SACCO, G. Somatic afferent transmission and cortical responsiveness during natural sleep and arousal in the cat. *Electroencephalography and Clinical Neurophysiology,* 1965, **18**, 354–368.

FEINBERG, I., KORESKO, R. L., GOTTLIEB, F., & WENDER, P. H. Sleep EEG and eye movement patterns in schizophrenic patients. *Comprehensive Psychiatry,* 1964, **5**, 44–53.

FISCHGOLD, H., SCHWARZ, B. A., & DREYFUS-BRISAC, C. Indicateur de l'état de presence et traces électroencéphalographiques dans le sommeil nembutalique. *Electroencephalography and Clinical Neurophysiology,* 1959, **11**, 23–33.

FISHER, C. Psychoanalytic implications of recent research on sleep and dreaming. I. Empirical findings. *Journal of the American Psychoanalytic Association,* 1965, **13**, 197–270.

FISHER, C., GROSS, J., & BYRNE, J. Dissociation of penile erection from REM period and rebound effect. Paper read at Association for the Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

FISHER, C., GROSS, J., & ZUCH, J. Cycle of penile erection synchronous with dreaming (REM) sleep. *Archives of General Psychiatry,* 1965, **12**, 29–45.

FISHER, S. Body image boundaries and hallucinations. In L. J. West (Ed.), *Hallucinations.* New York: Grune & Stratton, 1962. Pp. 249–260.

FISS, H., KLEIN, G. S., & BOKERT, E. Effects of dreaming and nondreaming sleep on fantasy in the waking state. *Archives of General Psychiatry,* 1966, **14**, 543–551.

FOULKES, D. Dream reports from different stages of sleep. *Journal of Abnormal and Social Psychology,* 1962, **65**, 14–25.

FOULKES, D. Theories of dream formation and recent studies of sleep consciousness. *Psychological Bulletin,* 1964, **62**, 236–247.

FREEDMAN, S. J., GRUNEBAUM, H. U., & GREENBLATT, M. Perceptual and cognitive changes in sensory deprivation. In P. Solomon, P. E. Kubzansky, P. H. Leiderman, J. H. Mendelson, R. Trumbull, &

D. Wexler (Eds.), *Sensory Deprivation.* Cambridge: Harvard University Press, 1961. Pp. 58–71.

GASTAUT, H., BATINI, C., BROUGHTON, R., FRESSY, J., ROGER, J., & TASSINARI, C. A. Étude électrographique des manifestations paroxystiques au cour du sommeil nocturne. *Revue Neurologique,* 1964, **110**, 309.

GASTAUT, H., & BROUGHTON, R. A clinical and polygraphic study of episodic phenomena during sleep. In J. Wortis (Ed.), *Recent advances in biological psychiatry.* Vol. 7. New York: Plenum Press, 1965. Pp. 197–221.

GELLHORN, E. *Physiological foundations of neurology and psychiatry.* Minneapolis: University of Minnesota Press, 1953.

GOODENOUGH, D. R., SHAPIRO, A., HOLDEN, M., & STEINSCHRIBER, L. A comparison of "dreamers" and "non-dreamers": Eye movements, EEGs, and the recall of dreams. *Journal of Abnormal and Social Psychology,* 1959, **59**, 295–302.

GOODENOUGH, D. R., LEWIS, H. B., SHAPIRO, A., JARET, L., & SLESER, I. Dream reporting following abrupt and gradual awakenings from different types of sleep. *Journal of Personality and Social Psychology,* 1965, **2**, 170–179.

GROSS, J., WEITZMAN, E. D., FISHER, C., & BYRNE, J. Penile erection in the monkey during REM periods. Paper read at Association for Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

HARTMANN, E. L. The D-State: A review and discussion of studies on the physiologic state concomitant with dreaming. *New England Journal of Medicine,* 1965, **273**, 30–35, 87–92.

HAWKINS, D. R., SCOTT, J., & THRASHER, G. Sleep patterns in enuretic children. Paper read at Association for the Psychophysiological Study of Sleep, Washington, March 1965.

HEAD, H. The conception of nervous and mental energy. II. Vigilance: A physiological state of the nervous system. *British Journal of Psychology,* 1923, **14**, 126–147.

HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

HENDLEY, C. D. Effects of drugs on sharp waves from the lateral geniculate nucleus and eye movements during low-voltage sleep in the cat. Paper read at Association for the Psychophysiological Study of Sleep, Brooklyn, March 1963.

HERNÁNDEZ-PEÓN, R. A cholinergic sleep system in the brain. Paper read at Department of Psychiatry, Maimonides Hospital, Brooklyn, October 9, 1964.

HERZ, A. Cortical and subcortical auditory evoked potentials during wakefulness and sleep in the cat. In K. Akert, C. Bally, & J. P. Schadé (Eds.), *Sleep mechanisms.* New York: Elsevier, 1965. Pp. 63–69.

HESS, R., JR. Sleep and sleep disturbances in the EEG. In K. Akert, C. Bally, & J. P. Schadé (Eds.), *Sleep mechanisms.* New York: Elsevier, 1965. Pp. 127–139.

HISHIKAWA, Y., & KANEKO, I. Electroencephalographic study on narcolepsy. *Electroencephalography and Clinical Neurophysiology,* 1965, **18**, 249–259.

HODES, R., & DEMENT, W. C. Depression of electrically induced reflexes ("H-reflexes") in man during low voltage EEG sleep. *Electroencephalography and Clinical Neurophysiology,* 1964, **17**, 617–629.

HODES, R., & SUZUKI, J. I. Comparative thresholds of cortex, vestibular system and reticular formation in wakefulness, sleep and rapid eye movement periods. *Electroencephalography and Clinical Neurophysiology,* 1965, **18**, 239–248.

JACOBSON, A., KALES, A., PAULSON, M., LEHMANN, D., & ZWEIZIG, J. R. Somnambulism: All-night EEG and psychological studies. Paper read at Association for the Psychophysiological Study of Sleep, Washington, March 1965. (a)

JACOBSON, A., KALES, A., PAULSON, M., LEHMANN, D., & ZWEIZIG, J. R. Somnambulism: All-night EEG and psychological studies. *Science,* 1965, **148**, 975–977. (b)

JAMES, W. *The principles of psychology.* New York: Henry Holt, 1890.

JOUVET, M. Paradoxical sleep—a study of its nature and mechanisms. In K. Akert, C. Bally, & J. P. Schadé (Eds.), *Sleep mechanisms.* New York: Elsevier, 1965. Pp. 20–62. (a)

JOUVET, M. Report on research in progress. Paper read at Association for the Psychophysiological Study of Sleep, Washington, March 1965. (b)

JOUVET, M., & JOUVET, D. A study of the neurophysiological mechanisms of dreaming. *Electroencephalography and Clinical Neurophysiology,* 1963, Supplement 24, 133–157.

JOUVET, M., & MICHEL, F. Délenchement de la "phase paradoxale" du sommeil par stimulation du tronc cérébral chez le chat intact et mesencéphalique chronique. *Comptes Rendus des Séance de la Société de la Biologie et de ses Filiales,* 1960, **154**, 636–641. (a)

JOUVET, M., & MICHEL, F. Nouvelles recherches sur les structures responsables de la "phase paradoxale" du sommeil. *Journal de Physiologie,* 1960, **52,** 130–131. (b)

KALES, A., JACOBSON, A., PAULSON, M. J., KALES, J. D., & WALTER, R. D. Somnambulism: Psychophysiological correlates. I. All-night EEG studies. *Archives of General Psychiatry,* 1966, **14,** 586–594.

KAMIYA, J. Behavioral, subjective and physiological aspects of drowsiness and sleep. In D. W. Fiske & S. R. Maddi (Eds.), *The functions of varied experience.* Hollywood, Ill.: Dorsey Press, 1961. Pp. 145–174.

KAMIYA, J. Behavioral and physiological concomitants of dreaming. Progress report to National Institute of Mental Health, 1962.

KANZOW, E., KRAUSE, D., & KÜHNEL, H. Die Vasomotorik der Hirnrinde in den Phasen desynchronisierter EEG-Activitat in naturlichen Schlaf der Katze. *Archiv für die gesamte Physiologie des Menschen und der Tiere,* 1962, **274,** 593–607.

KARACAN, I., & GOODENOUGH, D. R. REM deprivation in relation to erection cycle during sleep in young adults. Paper read at Association for the Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

KARACAN, I., & SNYDER, F. Erection cycle during sleep in *macaca mulatta* (preliminary report). Paper read at Association for the Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

KAWAMURA, H., & SAWYER, C. H. Differential temperature changes in the rabbit brain during slow wave and paradoxical sleep. Paper read at Association for the Psychophysiological Study of Sleep, Washington, March 1965.

KHAZAN, N., & SAWYER, C. H. "Rebound" recovery from deprivation of paradoxical sleep in the rabbit. *Proceedings of Society for Experimental Biological Medicine,* 1963, **114,** 536–539.

KLEIN, M. Etude polygraphique et phylogenetique des differents états de sommeil. Thèsis de Médecine, Lyon, Bosc. Edition, 1963.

KLEITMAN, N. *Sleep and wakefulness.* Chicago: University of Chicago Press, 1963.

LEVITT, R. Personal communication to the authors. Cited by W. Webb & H. W. Agnew. Sleep: Effects of a restricted regime. *Science,* 1965, **150,** 1745–1747.

LILLY, J. C. Mental effects of physical restraint and of the reduction of ordinary levels of physical stimuli on intact, healthy persons. *Psychiatric Research Reports,* No. 5, June, 1956.

MAGOUN, H. W. *The waking brain.* (2nd ed.) Springfield, Ill.: Charles C Thomas, 1963.

MANDELL, A. J., & MANDELL, M. P. Biochemical aspects of rapid eye movement sleep. *American Journal of Psychiatry,* 1965, **122,** 391–401.

MANDELL, M. P., MANDELL, A. J., & JACOBSON, A. Biochemical and neurophysiological studies of paradoxical sleep. In J. Wortis (Ed.), *Recent advances in biological psychiatry.* Vol. 7. New York: Plenum Press, 1965. Pp. 115–122.

MARON, L., RECHTSCHAFFEN, A., & WOLPERT, E. A. The sleep cycle during napping. *Archives of General Psychiatry,* 1964, **11,** 503–508.

MILLER, S. C. Ego-autonomy in sensory deprivation, isolation and stress. *International Journal of Psychoanalysis,* 1962, **43,** 1–20.

OMWAKE, K. T. Effect of varying periods of sleep on nervous stability. *Journal of Applied Psychology,* 1932, **16,** 623–632.

ORLINSKY, D. E. Psychodynamic and cognitive correlates of dream recall. Unpublished doctoral dissertation, University of Chicago, 1962.

OSWALD, I. *Sleeping and waking.* New York: Elsevier, 1962.

OSWALD, I., TAYLOR, A. M., & TREISMAN, M. Discriminative responses to stimulation during human sleep. *Brain,* 1960, **83,** 440–453.

PETRE-QUADENS, O. The ontogenesis of the phase of sleep corresponding to dreaming. *Journal of Neurological Sciences,* in press.

PORTNOFF, G., BAKELAND, F., GOODENOUGH, D. R., KARACAN, I., & SHAPIRO, A. Retention of verbal materials perceived immediately prior to onset of non-REM sleep. *Perceptual and Motor Skills,* 1966, **22,** 751–758.

RECHTSCHAFFEN, A., CORNWELL, P., & ZIMMERMAN, W. Brain temperature variations with paradoxical sleep in the cat. Read at Association for the Psychophysiological Study of Sleep, Washington, March 1965.

RECHTSCHAFFEN, A., & FOULKES, D. Effect of visual stimuli on dream content. *Perceptual and Motor Skills,* 1965, **20,** 1149–1160.

RECHTSCHAFFEN, A., HAURI, P., & ZEITLIN, M. Auditory awakening thresholds in REM and NREM sleep stages. *Perceptual and Motor Skills,* 1966, **22,** 927–942.

RECHTSCHAFFEN, A., & VERDONE, P. Amount of dreaming: Effect of incentive, adaptation to laboratory, and individual differences. *Perceptual and Motor Skills,* 1964, 19, 947–958.

RECHTSCHAFFEN, A., VERDONE, P., & WHEATON, J. Reports of mental activity during sleep. *Canadian Psychiatric Association Journal,* 1963, **8**, 409–414.

RECHTSCHAFFEN, A., WOLPERT, E., DEMENT, W. C., MITCHELL, S., & FISHER, C. Nocturnal sleep of narcoleptics. *Electroencephalography and Clinical Neurophysiology,* 1963, **15**, 599–609.

ROFFWARG, H. P., DEMENT, W. C., & FISHER, C. Preliminary observations of the sleep-dream pattern of neonates, infants, children and adults. In E. Harms (Ed.), *Problems of Sleep and Dream in Children,* Vol. II (International Series of Monographs on Child Psychiatry). New York: Pergamon Press, 1964. Pp. 60–72.

ROFFWARG, H. P., MUZIO, J. N., & DEMENT, W. C. Ontogenetic development of the human sleep-dream cycle. *Science,* 1966, **152**, 604–619.

ROLDAN, E., WEISS, T., & FIFKOVA, E. Excitability changes during the sleep cycle of the rat. *Electroencephalography and Clinical Neurophysiology,* 1963, **15**, 775–785.

ROSANSKI, J., & ROSEN, H. Musical hallucinations in ostosclerosis. *Confinia Neurologia,* 1952, **12**, 49–54.

ROSENSWEIG, N. Sensory deprivation and schizophrenia: Clinical and theoretical similarities. *American Journal of Psychiatry,* 1959, **116**, 326–329.

SAWYER, C. Feedback control of sexual behavior and pituitary-gonad function. Paper read at Interdepartmental Institute for Training in Research in the Behavioral and Neurologic Sciences, Albert Einstein College of Medicine, New York, April 1965.

SEARLES, H. F. The place of neutral therapist-responses in psychotherapy with the schizophrenic patient. *International Journal of Psychoanalysis,* 1963, **44**, 42–56.

SHIMAZONO, Y., HORIE, T., YANAGISAWA, Y., HORI, N., CHIKAZAWA, W., & SHOZUKA, K. The correlation of the rhythmic waves of the hippocampus with the behavior of dogs. *Neurologica Medico-Chirurgica,* 1960, **2**, 82–88.

SNYDER, F. The organismic state associated with dreaming. In N. S. Greenfield & W. C. Lewis (Eds.), *Psychoanalysis and Current Biological Thought.* Madison: University of Wisconsin Press, 1965. Pp. 275–315. (a)

SNYDER, F. Progress in the new biology of dreaming. *American Journal of Psychiatry,* 1965, **122**, 377–385. (b)

SNYDER, F. Toward an evolutionary theory of dreaming. *American Journal of Psychiatry,* 1966, **123**, 121–142.

SNYDER, F., HOBSON, A. J., MORRISON, D. F., & GOLDFRANK, F. Changes in respiration, heart rate and systolic blood pressure in relation to EEG patterns of human sleep. *Journal of Applied Physiology,* 1964, **19**, 417–422.

SOVRAD, D., & KAMAROVA, I. G. In preparation, cited by W. Webb & H. W. Agnew. Sleep: Effects of a restricted regime. *Science,* 1965, **150**, 1745–1747.

VERNON, J. A., MCGILL, T. E., GULICK, W. L., & CANDLAND, D. K. The effect of human isolation upon some perceptual and motor skills. In P. Solomon, P. E. Kubansky, P. H. Liederman, J. H. Mendelson, R. Trumbull, & D. Wexler (Eds.), *Sensory Deprivation.* Cambridge: Harvard University Press, 1961.

WEBB, W., & AGNEW, H. W. Sleep: Effects of a restricted regime. *Science,* 1965, **150**, 1745–1747.

WEBB, W., AGNEW, H. W., & WILLIAMS, R. Loose thoughts about fixed patterns. Paper read at Association for the Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

WEITZMAN, E., KRIPKE, D., POLLAK, C., & DOMINGUEZ, J. Cyclic activity in sleep of *macaca mulatta. Archives of Neurology,* 1965, **12**, 463–467.

WINTERS, W. D., & SPOONER, C. E. A neurophysiological comparison of gammahydroxybutyrate with pentobarbital in cats. *Electroencephalography and Clinical Neurophysiology,* 1965, **18**, 287–296.

WOOD, P. B. Dreaming and social isolation. Unpublished doctoral dissertation, University of North Carolina, 1962.

WURTZ, R. H. Steady potential shifts during arousal and deep sleep in the cat. *Electroencephalography and Clinical Neurophysiology,* 1965, **18**, 649–662.

ZARCONE, V., GULEVICH, G., & DEMENT, W. C. Some relationships between REM sleep and electrically induced convulsions. Paper read at Association for the Psychophysiological Study of Sleep, Gainesville, Fla., March 1966.

ZUNG, W. W. K., & WILSON, W. P. Response to auditory stimulation during sleep. *Archives of General Psychiatry,* 1961, **4**, 548–552.

ZUNG, W. W. K., & WILSON, W. P. Attention, discrimination and arousal during sleep. *Electroencephalography and Clinical Neurophysiology,* 1966, **20**, 623.

(Early publication received May 16, 1966)

Psychological Review
1966, Vol. 73, No. 6, 527–539

SOCIAL REINFORCEMENT EFFECTS AS A FUNCTION OF SOCIAL REINFORCEMENT HISTORY

REUBEN M. BARON[1]

Wayne State University

It was suggested that an individual's past history of social reinforcement importantly influences his present receptivity to social reinforcement. Affective and behavioral reactions to present levels of social reinforcement are assumed to be a complex function of both the direction and magnitude of any disparity that may exist between the present rate of social reinforcement and the individual's standard of social reinforcement (SRS). Disparities which produce negative affect (i.e., rates of social reinforcement which are either substantially above or below the SRS) are assumed, under certain specified conditions, to result in the individual varying his self-presentation (e.g., his task performance) in a direction which is likely to result in his being reinforced at a rate which he deems appropriate, i.e., a rate which either matches or moderately exceeds his SRS. It is thus hypothesized that the degree of conditionability found in a variety of studies concerned with the effects of social reinforcement on simple motor behaviors and verbal operants can be interpreted as involving attempts at mutual social control.

Williams (1964), in her review of the verbal conditioning literature, notes the increasing attention that has been paid in recent years to the effects of social reinforcement history on verbal conditioning. This research has primarily taken the form of an attempt to establish that a relationship exists between a given personality variable, such as anxiety or need for social approval (as defined operationally by scores on personality tests), and the conditioning of some verbal operant. This somewhat circuitous approach has led to a welter of contradictory and/or uninterpretable findings which at best merely ". . . furnish evidence of the complexity of the processes underlying conditionability . . . [Williams, 1964, p. 388]."[2] Lest the basic proposition at issue—the notion that an individual's past history of social reinforcement influences his present receptivity to social reinforcement—be obscured by such findings, it seems necessary to put less stress on the complex personality characteristics of the individual and more stress on the characteristics of the so-

---

[1] The author is indebted to J. W. Atkinson, E. E. Jones, and H. C. Kelman for helpful criticisms of an earlier version of this paper, but the opinions herein expressed are solely the responsibility of the author.

[2] The recent research efforts concerning the relationship between anxiety and verbal conditioning and need for social approval and verbal conditioning are illustrative of this confused, contradictory state of affairs. With regard to the relationship between anxiety and verbal conditioning, some studies have found better conditioning for high-anxious *S*s (Sarason, 1957; Taffel, 1955), some studies report significantly better conditioning for low-anxious *S*s (Buss & Gerjouy, 1958; Spielberger, DeNike, & Stein, 1965), and others report no relationship between anxiety and conditioning (Matarazzo, Saslow, & Pareis, 1960; Rowley & Keller, 1962). Similarly, while the work of Crowne and Marlow (1964) indicates that need for social approval facilitates the conditioning of verbal behavior, Spielberger, Berger, and Howard (1963) found no relationship between scores on the Crowne-Marlowe measure of need for social approval and the rate of verbal conditioning.

cial reinforcement history itself (e.g., the frequency, intensity, and variability of past social rewards).

Unfortunately, most investigators who have noted the importance of such parameters of the past history of social reinforcement see their effects almost solely in terms of creating differences in initial response strength upon which contemporary "shaping" procedures must be imposed (see, e.g., Staats & Staats, 1963). Such learning approaches typically fail to take into account the recent evidence on the social psychology of the psychological experiment gathered by investigators such as Riecken (1962), Orne (1962), and Rosenthal (1963), which indicates that situations such as those studied in the verbal conditioning paradigm represent ". . . an interpersonal event involving all the intricacies of *mutual* social control [Jones, 1964, p. 16]." For example, Riecken (1962) suggests that experimental outcomes are frequently more dependent on an implicit process of interpersonal bargaining between the subject (*S*) and the experimenter (*E*) than they are the result of formal task requirements; that is, *S* may or may not be willing to exchange doing what he perceives *E* wants him to do for a favorable evaluation.

The present conceptualization of the impact of the individual's history of social reinforcement on his present receptivity to social reinforcement suggests that a person's past history produces an internal norm or frame of reference which importantly influences the nature of the interaction between the subject and the reinforcing agent. Although any number of the major parameters of an individual's history of social reinforcement may be implicated in this process, for the purposes of the present exposition attention will be focused on the percentage or rate of past social reward as the major determinant of what will be referred to as an individual's *Social Reinforcement Standard* (SRS).

Conceptually, the notion of such a standard bears a strong resemblance to Helson's (1964) notion of an adaptation level (AL), and, indeed, Bevan (1963) has extended the notion of an AL to the area of nonsocial reinforcement. The crucial differences between the AL notion and the present concept of an SRS are as follows: (*a*) ALs have been restricted to the intensity dimension of stimulation; and (*b*) ALs have always been looked upon as neutral points or points of psychological indifference. The present notion of a social reinforcement standard, on the other hand, posits that the kind of internal referent that is established on the basis of past social reinforcement schedules is more than a neutral summary of one's past social reinforcement experiences; it is rather a preferred region around which one seeks to secure future social reinforcement. Such standards may be viewed as techniques of interpersonal uncertainty reduction. They allow us to smoothly coordinate our actions and interactions even with unknown others. Harvey (1965, p. 249) goes even further and suggests that a refutation of

> . . . one's standardized construal of a situation would be equivalent to a threat to the self, would portend negation of one's assumed means-end relations, and, if severe enough, would render the world chaotic and unstructured, destroying in the process one's means of fate control and techniques of defining and coping with his environment.

It may then be assumed that an individual's past history of social reinforcement defines for him a base line against which the adequacy or appropriateness of present social reinforcer inputs is judged. For the moment it will be further assumed that the level at which a person's social reinforcement thermostat is set will be isomorphic with the objective features of his past history of social reward. For example, it might be assumed that for an individual whose past frequency of social reward for achievement-oriented behavior was at a rate of approximately 66%, a region of reward ranging between 60% and 70% defines what an appropriate rate of social reward will be. The width of this range is assumed to vary with a number of factors, such as the variability which characterizes the individual's past history of social reward. For example, highly regular schedules may be assumed to produce rather narrow and rigid criteria as to what constitutes an acceptable rate of social reward.

Following Goffman (1959), it is further assumed that an individual does not merely define what he considers an appropriate rate of social reward and wait passively to see if the social environment will live up to these expectations. He communicates to others his notion of what an appropriate rate of social reward is for a given class of behaviors by means of his "self-presentation"; that is, by his selective display of certain attributes or behaviors in an attempt to influence the rate, direction, or type of social reward that he is to receive. The individual's use of various self-presentations may be seen as a defensive strategy designed to cope with potential, as well as actual, disparities which may occur between his social reinforcement standard and the contemporary rate of social reinforcement. Thus, the attributes selected for display are likely to be ones which will clearly influence the favorability of the impression *S* is making. That is, the individual is assumed to display those personal attributes which are *indicative of his ability to satisfy the needs of a particular other.* For example, let us assume that most *S*s enter a typical learning experiment with a social reward standard that calls for a moderately high rate of social reward for achievement-related behaviors. Furthermore, it may be assumed that *E*s who are college instructors are particularly likely to give the impression, whether they intend to or not, that they expect and value rapid learning (see Rosenthal, 1963, for a discussion of *E*'s hypotheses as unintended determinants of experimental results). In such a situation, to the extent that *S* perceives that *E* wants to think of himself as a good instructor, a rapid rate of learning or conditioning may be interpreted as a self-presentation strategy designed to please *E*, and thereby insure *S* his desired rate of social approval.

It is assumed that a self-presentation is particularly likely to take the form of variations in task performance when the difficulty of the task set for *S* is somewhat below his skill or capacity, since it is under these conditions that the entire range of performance from excellent to poor is under *S*'s voluntary control (see Riecken, 1962, for an elaboration of this point). It may also be noted that the work of Jones, Gergen, Gumpert, and Thibaut (1965) indicates that *S*s are significantly more likely to display attributes (e.g., traits and opinions) which *E* indicates he values when the standards for task evaluations are subjective rather than objective.[3]

Typically, laboratory studies of social reinforcer effects require *S*s to make verbal or sensory-motor discriminations which are rather simple. Furthermore, *S* is usually confronted with a task for which the criteria for whether he has made a right or wrong response are left rather vague. Under such conditions, it seems likely that at least part of the response variance can be accounted for by *S*'s use of the "arts of impression management" to insure a rate of social reinforcement which he deems appropriate, given his past history of social reinforcement.

---

[3] In Jones et al. (1965) the openness or closedness of the context of performance evaluation was varied orthogonally. In the open or subjective-evaluation condition it was made clear to *S*s that there were no specific criteria for correctness other than the supervisor's judgment, while in the closed or objective-evaluation condition *S*s were informed that the supervisor would have a list of the correct answers in front of him as he evaluated their solutions.

DISCREPANCIES FROM THE REINFORCEMENT BASE LINE

Investing the notion of a social reinforcement standard with so much motivational relevance leads to the following hypothesis: *Substantial disparities above, as well as below,* the person's SRS will produce a state of negative affective arousal with a concomitant activation of various self-presentations designed to eliminate such disparities. Whether small or moderate departures from this base line are also likely to produce negative affect, regardless of their direction, is a more complicated issue. The "butterfly curve" hypothesis of McClelland, Atkinson, Clark, and Lowell (1953) and the experimental investigations of Haber (1958) in the realm of aesthetics, as well as simple sensory experiences, suggest that small discrepancies from a base line in either direction produce positive affect, while larger discrepancies produce negative affect. Further, stimulus inputs at the base-line level are assumed to produce the negative affect of boredom.

Two kinds of factors weigh against our generalizing these notions to the functioning of a social reinforcement base line. First, it hardly seems plausible that when the stimulus series involves social rewards, information which confirms the accuracy of our construal of a social situation will result in the negative affect of boredom. Second, the butterfly-curve hypothesis implies that it is *only* the magnitude of discrepancy that is crucial, not the direction. Contrary to this assumption, Harvey and Clapp (1965) have experimentally demonstrated the importance of taking directionality into account, at least when the stimulus series has implications for the self-esteem of *S*. Their work indicates that even small negative discrepancies from a person's self-construal may produce negative affect, while positive discrepancies of even moderate size produce positive affect. Since social reinforcement input obviously has relevance to the self-esteem of *S*, it seems necessary to take into account the directionality of the discrepancy, at least when dealing with small- or moderate-size disparities. However, as we will attempt to demonstrate below, it will be assumed that large disparities from a social reinforcement base line, regardless of their direction, produce negative affect.

AN EXCURSION INTO THE PSYCHOLOGY OF OVERREWARD

The notion that, where the disparity from the social reward standard is large, even positive discrepancies will produce negative affect, and, hence, activate a self-presentation strategy designed to lower the level of social approval and/or increase the level of social disapproval, seems both counterintuitive and contrary to the position that most "game-theoretic" or "exchange" oriented investigators are likely to take.[4] Therefore, we will devote considerable pains at this point to buttress this derivation.

---

[4] For example, Thibaut and Kelley (1959), in elaborating their "exchange" model of social interaction, assume that the expected level of social reward does not become the preferred or optimal level of reward—the individual always seeks a relationship which is better than the average relationship. The

The case for such a derivation may be made as follows: First, it may be argued that outcomes which are substantially above one's previous reinforcement experience may produce negative affect, and hence may be avoided because they commit an individual to a public image which he realizes he will be unable to sustain. That is, greater than expected reward levels may produce negative affect because the person anticipates future disappointment and frustration for failing to be able to perform at a level commensurate with this higher than previously attained level of reward. Furthermore, greater than expected reward levels may produce guilt feelings in the person for having failed to achieve such reward levels up until now. Harvey and Clapp's (1965) and Van Ostrand's (1960) finding that persons of low self-esteem are less able than persons of high self-esteem to accept favorable evaluations from others may be interpreted in these terms.

Another class of factors which may cause rewards which are greater than those previously encountered to produce psychological discomfort is assumed to stem from considerations of equity and "distributive justice" (see Adams, 1963; Homans, 1961). These conceptions reflect the following kinds of phenomena. The person feels that the amount of effort he has expended or the quality of the work he has performed does not merit an increased level of reward because relevant positive-reference persons whose performances are similar in quality and involve equal effort are being rewarded at a lower rate. It may also be noted that greater than previously encountered reward levels may be rejected because they raise questions as to the credibility of the reinforcing agent—they raise suspicions as to his possible ulterior motives. The recent critiques of dissonance theory offered by Rosenberg (1965) and Janis and Gilmore (1965) underscore this point. Moreover, Jones, Jones, and Gergen (1963) report that a too ready reinforcement of the correctness of another's opinions can, under certain conditions, result in an increase in negative affect toward the overly compliant person.

AREAS OF APPLICATION

Lest the reader be misled, it should be noted that what follows is not intended to be an unbiased or complete review of various areas in which social reinforcement effects have been studied. The major purpose of such a review is not to establish the validity of our approach; this can only be accomplished by a separate program of research which is specifically tied to our concep-

---

more any particular relationship exceeds the expected level of reward, the better the individual is assumed to like it. In addition to arguments advanced in the text of this paper, the following may be noted. Thibaut and Kelley mainly focus on social situations involving voluntary associations. In the realm of voluntary associations it may be assumed that individuals do not ordinarily seek out persons who are likely to flatter them much beyond their own conceptions of their worth, just as they are likely to avoid interacting with persons who habitually insult them. The range of outcomes, then, to which a person is ordinarily exposed is a preselected one. Within such a relatively *restricted* band of outcomes it may indeed turn out that people always prefer the most favorable reward level possible. Within the laboratory social reinforcement situation, as well as in certain real-life situations involving involuntary association (e.g., in a formal work hierarchy), the individual may be forced to face a range of reward which is considerably beyond that which he would tolerate in an informal voluntary association. Thus, Thibaut and Kelley (1959) may err in the sense of overgeneralizing the principle of a nonselective preference for higher rewards to *all* interpersonal encounters, voluntary and otherwise.

tualization. The major purpose of such a selective sampling is rather to provide examples of the kinds of situations for which the present notions might be profitably taken into account.

Studies Involving Negative Discrepancies

If normal children may be assumed to enter an experiment with a history of a fairly high rate of positive social reinforcement for well-behaved interactions with adults, we may expect that for such children preexperimental encounters involving isolation, nonsupport, or negative reinforcement will create negative affect, and hence trigger self-presentations designed to raise the rate of social approval being received. Since all of the studies to be cited involve learning tasks that are of a simple, nonchallenging nature, it will be argued that these self-presentations are likely to take the form of task behaviors which are likely to result in an increment in social reinforcement; that is, it is expected that high rate of conditioning will be found in a social reinforcement situation which immediately follows the under-reward experience.

Thus, the increased rate of conditioning to a social reinforcer following a period of social isolation found by Gewirtz and Baer (1958a, 1958b), Erickson (1962), and Lewis and Richman (1964) all fit the present view of the processes underlying social reinforcer effectiveness in certain specified situations. The relevance of the present approach is also supported by (*a*) the superior learning found by Hartup (1958) following a sudden withdrawal of a nurturant relationship, (*b*) the increments in the incentive valve of a social reward following a negative reinforcement found by Lewis and Richman (1964) and (*c*) the finding of Walters and Ray (1960) that, for a child, being left alone with an adult stranger who ignores him increases the effectiveness of subsequent social reward. Cohen, Greenbaum, and Mansson's (1963) finding that, even with college students, the sudden withdrawal of social reinforcement after an initial high rate of approval results in a more rapid rate of verbal conditioning in a subsequent reinforcement situation than occurs for *S*s who did not experience such a negative discrepancy also fits the present assumptions.

Additional evidence in support of the present approach may be found in a study by Gilmore and Zigler (1964) involving the effects of social reinforcement and nonreinforcement on the task persistence of first- and latter-born children. Gilmore and Zigler's assumption that first-born children have had past social reinforcement experiences involving a high rate of social approval makes their findings particularly relevant. These investigators found that first-born *S*s experienced a good deal more discomfort under nonreinforcement than positive reinforcement and manifested behavior which can be interpreted as a strategy to *raise* the rate of positive reinforcement they were receiving; that is, they showed significantly more persistence on a boring task in a nonreinforcement than in a reinforcement treatment. This persistence may be seen as a tactic of self-presentation used by the first-borns to inform *E* that they are good, conscientious *S*s, and therefore should be given more social approval.

The author's supposition that in the above kinds of task situations *S*'s attempt to validate his SRS may be a more important determinant of his task performance than any problem-solving set which is part of the "demand characteristics" of the conditioning situation is directly supported by the following observations of Lewis and Richman (1964). Working with third-grade elementary school children, they found that *S*s who had preexperimental experiences of social isolation or negative reinforcement appeared to ". . . be *less* interested in a solution to the problem, and *more* interested in obtaining social reinforcement [p. 256]."

Positive Reward Discrepancies

Some studies that may be interpreted as involving positive *S*'s SRS will now be examined. A number of investigators have found that schizophrenics experience discomfort and do not condition or perform well when positive-reinforcement procedures are used.[5] A recent study by Berkowitz (1964) is particularly striking in this connection:

> Reaction times of 39 chronic schizophrenics and 39 normal adults were tested with and without reinforcement after exposure to 1 of 3 conditions of contact with *E*: (*a*) no contact, (*b*) contact with an unresponsive and aloof *E*, and (*c*) contact with a warm and accepting *E* [p. 522].

Contrary to Berkowitz's original hypothesis, the schizophrenics in the warm condition had significantly *slower* reaction times than those in the aloof and no-contact conditions. A number of studies of verbal conditioning have found that schizophrenics show better conditioning when nonreinforcement or negative reinforcement is used (e.g., Cohen & Cohen, 1960; Leventhal, 1959). Similarly, Atkinson and Robinson (1961) and Koppenhaver (1961) compared verbal reward and punishment in a paired-associates learning task and found that punishment improved schizophrenic learning relative to that obtained with reward.

If it may be assumed that schizophrenics have had a premorbid social reinforcement history characterized by nonreinforcement and/or negative reinforcement for a variety of behaviors, then the above results may be interpreted as follows: Schizophrenics evaluate positive social reinforcement as an inappropriate state of affairs. Thus, the poor learning or slow rate of conditioning found with schizophrenics when positive reinforcement was used may be seen as a defensive strategy designed to compel the reinforcing agent to produce the kinds of reinforcer events with which schizophrenics are likely to feel comfortable. That is, the schizophrenic who fails to "learn" under conditions of positive reinforcement may be seen as utilizing a self-presentation designed to convey the impression that he is stupid and inept, a presentation of self which is likely to elicit the nonreinforcement and/or negative reinforcement necessary to validate his SRS.

Evidence may also be adduced to show that even "normal" *S*s find substantial positive discrepancies disturbing. In a recent study Marlowe (1962) found that after 10 minutes of receiving 100% reinforcement for making positive self-references his *S*s showed a *significant decline* in the rate of emission of such utterances, with one *S* spontaneously commenting, "it's awfully hard to sit here and toot your own horn [cited in Crowne & Marlowe, 1964, p. 61]." This lowered rate of conditioning may be seen as a strategy adopted by *S*s to force *E* to lower the rate at which he was rewarding positive self-references, because after a while *S*s view the high rate of reward as inappropriate. Further, an experiment by Aronson and Carlsmith (1962) demonstrates that even when *S*'s reinforcement history is defined in terms of negative-reinforcement encounters experienced in a laboratory situation, a sudden confrontation with a highly rewarding state of affairs produces discomfort and indeed, when given the opportunity, positive-discrepancy *S*s try to lower the level of their performance score.

---

[5] Some investigators, notably Rodnick and Garmezy (1957) and their associates, have suggested that social censure generally disrupts rather than improves the performance of schizophrenics. The majority of findings in this area, however, do not support the position of Rodnick and Garmezy. In this connection a recent review article by Buss and Lang (1965) may be cited. They point out that, taken together, studies in this area ". . . do not support the hypothesis that the performance of schizophrenics is disrupted by stimuli connoting censure . . . [p. 5]." According to Buss and Lang (1965) earlier positive results have either been contradicted by later results, only weakly corroborated by later results, or are open to alternate interpretations because of methodological problems. Furthermore, as has been suggested above in the text of this paper, a number of studies have demonstrated that ". . . mild verbal censure of the type used by Garmezy and Rodnick and their students *facilitated* the performance of schizophrenics [Buss & Lang, 1965, p. 11]."

Finally, the present author, in an exploratory study concerned with the effects of degree and direction of reward discrepancy, found that for *S*s who initially had been reinforced at a 33% rate for a particular class of emotional labeling responses to a facial photograph a switch to a 66% rate of social reinforcement produced a more positive state of affect and a *higher* rate of emission of the reinforced response class than did a switch to a 100% rate of social reinforcement.

OPERATIONALIZING THE SRS

Before closing the present formulation, a brief discussion of possible means of operationalizing the notion of a social reinforcement standard seems in order. Two widely divergent approaches may be used. The first involves the development of various techniques to *assess* *S*s' past histories of social reinforcement; the other involves generating a set of experimental paradigms for creating in the laboratory histories of social reinforcement for behaviors whose occurrence is not tied to any clear expectation of social reward or punishment. Some examples of each of these approaches will be presented, and then a paradigm developed by the author will be described in some detail.

Any of the many approaches that have been used to assess the prior history of social reinforcement of *S*s could be used to infer their SRS given the assumption of an isomorphism between the SRS and the objective properties of *S*s' actual reinforcement histories. For example, a number of investigators have constructed history-of-reinforcement questionnaires which describe behavioral situations and then ask *S* to choose from a list of parental reaction such as kisses, isolation, deprivation of privileges, etc. A variant on this approach asks *S* to infer how a particular parent felt when *S* engaged in various behaviors (see, e.g., Crandall, 1963). Other investigators have used objective biographic data. Zigler (1961) inferred degree of prior deprivation of social reinforcement by having judges do a blind coding of detailed biographic data obtained about *S*'s early life; for example, information was gathered on whether *S*'s parents were separated when he was a child, if the parents were ever institutionalized, separated, divorced, etc. Similarly, Crandall, Good, and Crandall (1964) used school records as indexes of the child's probable history of reinforcement in everyday achievement experiences.

Since Crandall (1963) found a significant positive correlation between *S*s' generalized expectancies of success in her task situation and her questionnaire measure of social reinforcement, and Crandall et al. (1964) found significant positive correlations between the child's actual performance in school and his generalized expectancy of success, it seems that the assumption that there is an isomorphism between *S*'s actual reinforcement history and his standardized construal of that history is a reasonable one. Furthermore, since Crandall (1963) found *S*'s Generalized Expectancy (GE) a better predictor of present social reinforcement effectiveness than her questionnaire measure, it may be possible to infer *S*'s SRS by obtaining expectancy estimates prior to the performance of a given task. For example, *S*s might be asked to estimate the probability of their receiving social approval or disapproval for a given behavior in a variety of social reinforcement contexts such as with peers and high- and low-status others.[6]

While the assessment approach complemented with a measure of *S*'s GE provides a promising start to finding a valid set of procedures to operationalize *S*'s SRS, it should be noted that the correlations described above, while significant, were sufficiently small (between 21 and 32) to leave room for a good deal of unaccounted-for variance. It is therefore suggested that greater variance control might be obtained if expectancy estimates were obtained in conjunction with the creation of a social reinforcement history de novo in the laboratory. A general paradigm designed to accomplish this has been developed by the author and is described below.

EXPERIMENT I

Method

A verbal conditioning task was used which required that *S* judge the emotional state of a person from a facial photograph. *S* selected an emotional label from a list of 6—love, joy, contentment, anger, fear, disgust. The labels represented two general response classes: positive and negative emotions. The task was presented to *S* as having a twofold purpose:

1. First, the task was described as an attempt to validate a universal list of emotional categories developed by the experimenter. (This instruction was designed to heighten the level of mutual interdependence that exists in the *E-S* relationship.) 2. Second, *S* was told that performance on this task is a good measure of social sensitivity in the real world, and that his responses would be compared with norms obtained at the university. (This instruction was designed to increase the ego involvement of *S*.) The pictures used were taken from a past Wayne State University yearbook and were selected so as to be neutral in expression; that is, pictures having obvious cues as to the emotional state were eliminated. (This was done to break down any possible strong stimulus binding and therefore to make *S*'s judgments particularly susceptible to influence by *E*'s reinforcements.)

The following experimental procedure was then instituted. An operant series consisting of 15 nonreinforced trials was run. During this series, *E* showed *S*s 15 pictures, one at a time, giving them 10 seconds to make a judgment. The response class (i.e., positive or negative emotional labels) that was least used during the operant series was designated the "correct" one for purposes of future reinforcement. After the operant series was completed, and before the reinforcement treatments were given, *E* obtained *S*s' estimates of how well they expected to do on this task (GE measure).

An *acquisition series* or the building up of a social reinforcement history period followed the operant trials. *S*s were randomly assigned to either a 100% or a 33% fixed-ratio contingent-reinforcement condition. The *acquisition series* involved 60 trials—4 blocks of 15 trials at either of the two base-line rates. The administration of social reinforcement was accomplished by *E* saying "that's good" or "that's fine" for the correct response on one or the other of the above schedules. After the acquisition series was completed, *S*s were asked to estimate (*a*)

---

[6] The term Generalized Expectancy (GE) can be traced to Rotter (1954). He defines GE as "the generalization of the expectancies for the same or similar reinforcements to occur in a present situation as occurred in past situations for the same, or functionally related behaviors [p. 166]." Rotter goes on to suggest that the GE effect will weigh more heavily in novel situations than in those in which *S* has had a series of experiences. It is this aspect of the GE concept that makes it particularly useful for our purposes.

TABLE 1

Shift Series

Acquisition rate	100%	66%	33%
100%			
33%			

the proportion of judgments on which they thought they actually were correct, and (*b*) the proportion they expected to get correct on future trials.

The acquisition series was followed by a *shift series* involving 4 more blocks of 15 trials each. During the shift phase, each of the original base-line groups was broken into 3 subgroups generating the 2 × 3 factorial design shown in Table 1.

Following the shift series, *S*s were once more asked to estimate how well they actually thought they did and how well they expected to do on future trials. An extinction series consisting of 4 blocks of 15 nonreinforced trials was then instituted. After the extinction series was completed, *S*s were asked to make a final estimate of how well they thought they actually did. (A control group involving *S*s making judgments for 180 trials without any reinforcement was not included in the original pilot study, but is being run as part of a larger replication.)

The conditioning phase was followed by an assessment period during which questionnaire and interview data were collected. Information was gathered concerning *S*'s affective state during the experiment, periods of possible shifts in mood, and his general and specific evaluation of the experiment, *E*, and the task. (For the most part a semantic-differential format involving lists of bipolar adjectives was used to obtain this information.) Data were also collected as to *S*'s perception of how ego involved he was, how much he wanted to obtain *E*'s approval, and his notions about the purposes of the experiment. A detailed "awareness interview" patterned after those used by Dulany (1962) and Spielberger (1962) completed the experiment.[7]

Summary and Conclusions

The present attempt to conceptualize the interpersonal processes underlying the postulated relationship between *S*'s prior history of social reinforcement and his present receptivity to social reinforcement may be summarized as follows:

1. The individual's past social reinforcement experiences produce an internal norm or standard against which the adequacy of present social reinforcements is judged.

2. The individual's construal of such experiences (i.e., his standard of social reinforcement) is assumed to be essentially isomorphic with the objective properties of his past history of social reinforcement.

3. The affective consequences of an individual's reactions to present social reinforcement are assumed to be a complex function of both the direction and magnitude of any disparity that may exist between the present rate of social reinforcement and the individual's SRS.

4. Disparities which produce negative affect (i.e., rates of social reinforcement which are either sub-

---

[7] Although a sample of only 36 female subjects (6 in a cell) was used in the pilot study, some interesting, if tentative, trends emerged. First, for affective arousal and rate of conditioning it is clear that both the direction and magnitude of discrepancy from a reinforcement base line must be taken into account. Thus, while both conditions of moderate discrepancy (100%–66%, 33%–66%) produced less negative affect than the conditions of substantial discrepancy (100%–33%, 33%–100%), in every case the negative-discrepancy conditions produced more negative affect than a positive discrepancy of identical magnitude. The conditioning data yielded similar results. The conditions of moderate discrepancy (100%–66%, 33%–66%) each produced a higher rate of conditioning than the conditions of substantial discrepancy (100%–33%, 33%–100%). Furthermore, as with the affect data, each of the positive-discrepancy treatments produced a higher rate of conditioning than did the negative-discrepancy treatments of equal magnitude. A more detailed report of the data yielded by this paradigm will be deferred until the present replication, which is utilizing a larger sample, is completed.

stantially above or below the SRS) are assumed to result in the individual's varying his self-presentation in such a way as to induce the reinforcing agent to reward him at a rate which he construes as appropriate, that is, at a rate which either matches or moderately exceeds his SRS.

5. When the context of evaluation is vague, and the learning task is of a simple, nonchallenging nature, these self-presentations may take the form of variations in task performance whose direction is determined by *S*'s assessment of what performances are likely to allow him to gain control over the social reward responsivity of the reinforcing agent. Task performance in such situations may be viewed as a defensive strategy designed to cajole *E* into reinforcing *S* at a rate of social reward with which he feels comfortable.

It is hoped that by adopting the perspective of *S*, we have made some case for the value of treating social reinforcer effects as the result of a social interaction between two equally involved, if not equally powerful, actors.

REFERENCES

Adams, J. S. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 1963, **67**, 422–436.

Aronson, E., & Carlsmith, J. M. Performance expectancy as a determinant of actual performance. *Journal of Abnormal and Social Psychology*, 1962, **65**, 178–182.

Atkinson, R. L., & Robinson, N. M. Paired-associate learning by schizophrenics and normal subjects under conditions of personal and impersonal reward and punishment. *Journal of Abnormal and Social Psychology*, 1961, **62**, 322–326.

Berkowitz, H. Effects of prior experimenter-subject relationships on reinforced reaction times of schizophrenics and normals. *Journal of Abnormal and Social Psychology*, 1964, **69**, 522–531.

Bevan, W. The pooling mechanism and the phenomena of reinforcement. In O. J. Harvey (Ed.), *Motivation and social interaction—cognitive determinants.* New York: Ronald, 1963. Pp. 18–34.

Buss, A. H., & Gerjouy, I. R. Verbal conditioning and anxiety. *Journal of Abnormal and Social Psychology*, 1958, **57**, 249–250.

Buss, A. H., & Lang, P. J. Psychological deficit in schizophrenia: I. Affect, reinforcement, and concept attainment. *Journal of Abnormal Psychology*, 1965, **70**, 2–24.

Cohen, E., & Cohen, B. Verbal reinforcement in schizophrenia. *Journal of Abnormal and Social Psychology*, 1960, **60**, 443–446.

Cohen, A. R., Greenbaum, C. W., & Mansson, H. H. Commitment to social deprivation and verbal conditioning. *Journal of Abnormal and Social Psychology*, 1963, **67**, 410–421.

Crandall, V. Reinforcement effects of adult reactions and non-reactions on children's achievement expectations. *Child Development*, 1963, **34**, 335–354.

Crandall, V., Good, S., & Crandall, V. J. Reinforcement effects of adult reactions and non-reactions on children's achievement expectations: A replication study. *Child Development*, 1964, **35**, 485–497.

Crowne, D. P., & Marlowe, M. *The approval motive: Studies in evaluative dependence.* New York: Wiley, 1964.

Dulany, D. E., Jr. The place of hypotheses and intentions: Analysis of verbal control in verbal conditioning. In C. W. Erikson (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962. Pp. 102–120.

Erickson, M. T. Effects of social deprivation and satiation on verbal conditioning in children. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 953–957.

Gewirtz, J. L., & Baer, D. M. Deprivation and satiation of social reinforcers as drive condition. *Journal of Abnormal and Social Psychology*, 1958, **57**, 165–172. (a)

Gewirtz, J. L., & Baer, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *Journal of Abnormal and Social Psychology*, 1958, **56**, 49–56. (b)

Gilmore, J., & Zigler, E. Birth order and reinforcer effectiveness. *Child Development*, 1964, **35**, 193–200.

Goffman, E. *The presentation of self in everyday life.* New York: Doubleday, 1959.

Haber, R. N. Discrepancy from adaptation level as a source of affect. *Journal of*

Experimental Psychology, 1958, **56**, 370–375.

HARTUP, W. W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of pre-school children. *Child Development*, 1958, **13**, 191–201.

HARVEY, O. J. Cognitive aspects of affective arousal. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition, and personality*. New York: Springer, 1965. Pp. 242–263.

HARVEY, O. J., & CLAPP, W. F. Hope, expectancy, and reactions to the unexpected. *Journal of Personality and Social Psychology*, 1965, **2**, 45–52.

HELSON, H. *Adaptation level theory*. New York: Harper, 1964.

HOMANS, G. C. *Social behavior: Its elementary forms*. New York: Harcourt, Brace & World, 1961.

JANIS, I. L., & GILMORE, B. J. The influence of incentive conditions on the success of role playing on modifying attitudes. *Journal of Personality and Social Psychology*, 1965, **1**, 17–27.

JONES, E. E. *The psychology of ingratiation*. New York: Appleton-Century-Crofts, 1964.

JONES, E. E., GERGEN, K. J., GUMPERT, P., & THIBAUT, J. W. Some conditions affecting the use of ingratiation to influence performance evaluation. *Journal of Personality and Social Psychology*, 1965, **1**, 613–625.

JONES, E. E., JONES, R. G., & GERGEN, K. J. Some conditions affecting the evaluations of a conformist. *Journal of Personality*, 1963, **31**, 270–288.

KOPPENHAVER, N. D. The effects of verbal and non-verbal reinforcement on the performance of schizophrenic subjects. Unpublished doctoral dissertation, Purdue University, 1961.

LEVENTHAL, A. M. The effects of diagnostic category and reinforcer on learning without awareness. *Journal of Abnormal and Social Psychology*, 1959, **59**, 162–166.

LEWIS, M. E., & RICHMAN, S. Social encounters and their effect on subsequent social reinforcement. *Journal of Abnormal and Social Psychology*, 1964, **69**, 253–257.

MARLOWE, D. Need for social approval and the operant conditioning of meaningful verbal behavior. *Journal of Consulting Psychology*, 1962, **26**, 79–83.

MATARAZZO, J. D., SASLOW, G., & PAREIS, E. N. Verbal conditioning of two response classes: Some method of logical considerations. *Journal of Abnormal and Social Psychology*, 1960, **61**, 190–206.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, A. R., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776–783.

RIECKEN, H. A program for research on experiments in social psychology. In N. F. Washbourne (Ed.), *Decisions, values, and groups*. Vol. 2. London: Pergamon Press, 1962. Pp. 75–141.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: University of Nebraska Press, 1957. Pp. 109–184.

ROSENBERG, M. J. When dissonance fails: On eliminating evaluative apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1965, **1**, 28–42.

ROSENTHAL, R. On the social psychology of the psychological experiment. *American Scientist*, 1963, **51**, 268–283.

ROTTER, J. B. *Social learning and clinical psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1954.

ROWLEY, V., & KELLER, E. D. Changes in children's verbal behavior as a function of social approval and manifest anxiety. *Journal of Abnormal and Social Psychology*, 1962, **65**, 53–57.

SARASON, I. G. Effects of anxiety and two kinds of motivating instructions on verbal learning. *Journal of Abnormal and Social Psychology*, 1957, **54**, 166–171.

SPIELBERGER, C. D. Awareness in verbal conditioning. In C. W. Ericksen (Ed.), *Behavior and awareness*. Durham: Duke University Press, 1962. Pp. 73–101.

SPIELBERGER, C. D., BERGER, A., & HOWARD, K. Conditioning of verbal behavior as a function of awareness, need for social approval, and motivation to receive reinforcement. *Journal of Abnormal and Social Psychology*, 1963, **67**, 241–246.

SPIELBERGER, C. D., DENIKE, L. D., & STEIN, L. S. Anxiety and verbal conditioning. *Journal of Personality and Social Psychology*, 1965, **1**, 229–239.

STAATS, A. W., & STAATS, C. K. *Complex human behavior*. New York: Holt, Rinehart & Winston, 1963.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology,* 1955, **51**, 496-501.

THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups.* New York: Wiley, 1959.

VAN OSTRAND, D. Reactions to positive and negative information about the self as a function of certain personality characteristics of the recipient. Unpublished master's thesis, University of Colorado, 1960.

WALTERS, R. H., & RAY, W. Anxiety, social isolation and reinforcer effectiveness. *Journal of Personality,* 1960, **28**, 358-367.

WILLIAMS, J. H. Conditioning of verbalization: A review. *Psychological Bulletin,* 1964, **62**, 383-393.

ZIGLER, E. Social deprivation and rigidity in the performance of feebleminded children. *Journal of Abnormal and Social Psychology,* 1961, **62**, 413-421.

(Received September 2, 1965)

Psychological Review
1966, Vol. 73, No. 6, 540–549

AN ANALYSIS OF SOME SHORTCOMINGS IN THE INTERFERENCE THEORY OF FORGETTING

BENTON J. UNDERWOOD AND BRUCE R. EKSTRAND

Northwestern University

Tests of interference theory of forgetting as applied to extraexperimental sources of interference have been generally negative. The hypothesis was tested that the negative findings result from the fact that linguistic associations, presumed to interfere, do not do so because of their high level of learning, the distribution of their learning over time, or both. The A-B, A-C paradigm was employed, with A-B being varied from 12–80 trials, and was learned either by massed practice or by distribution of practice over 4 days. Recall of A-C after 24 hrs. showed that distribution of A-B resulted in only small proactive inhibition (PI), while massing resulted in heavy PI. Degree of A-B learning influenced the recall only under massed A-B learning. Some implications for the development of forgetting theory were presented.

The orderly development of a detailed interference theory of forgetting has been disrupted in a most disorderly but convincing manner: The data are not supporting certain predictions from the theory. The failures are centered primarily in the theory as extended to account for forgetting from extraexperimental sources of interference (Underwood & Postman, 1960), and these failures are responsible for the present paper. A hypothesis to account for the failures has been developed, and an experimental test has been made of it. Some background is necessary in order to understand the origin of the hypothesis; more particularly, it seems necessary to state explicitly how the interference theory has been extended to incorporate extraexperimental sources of interference.

The general theory that forgetting results from interference among associations was developed primarily as a result of studies of retroactive inhibition (RI), in which the interference originated from interpolated associations. In relatively recent years the discovery of proactive inhibition (PI) required extension of sources of interference to include associations learned prior to the learning of the task to be recalled. From the results of studies of RI and PI, it seemed necessary to add additional mechanisms to account for differences in forgetting observed under various conditions, thereby accounting for variations in the magnitude of the interference. Two mechanisms were added, namely, extinction of associations when successive lists involved conflicting associations (as in A-B, A-C), and the spontaneous recovery over time of extinguished associations. Some such mechanisms seemed necessary to account for certain results. For example, the increase in PI with increases in the length of the retention interval could be handled by the recovery of extinguished associations in the first list which interfered increasingly with the recall of second-list associations.

The extension to account for forgetting from extraexperimental sources of

---

[1] This work was supported by Contract Nonr-1228 (15), Project NR 154-057, between Northwestern University and the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose by the United States Government.

interference was done with complete fidelity to the theory as developed in the formal RI and PI paradigms. When a naïve subject (*S*) learned a single list, it was presumed that the associations required conflicted to some extent with linguistic associations learned previously outside the laboratory. Therefore, learning of the list was assumed to produce some extinction of the preexperimental associations which, with the passage of time, however, recovered to produce interference (PI) at recall. The role of RI was minimized on the grounds that acquisition of conflicting associations during the retention interval was much less probable than having already acquired such associations prior to learning the list.

The theory further specified what conditions should produce variations in the amount of PI. Again, the translation was made from the formal RI and PI situation. The greater the conflict was between associations, the greater the RI and PI (e.g., A-B, A-C versus A-B, C-D). Variations in number and strength of conflicting associations were spelled out in terms of unit-sequence interference and letter-sequence interference, which in turn led to specific predictions concerning differential rates of forgetting of lists constructed to minimize or maximize interference with linguistic habits.

It is not necessary to recite the rather agonizingly long list of experiments using different materials and different types of tasks, all of which seemed to fit rather precisely the specifications of the theory, but which failed to show the predicted differences in retention (e.g., Ekstrand & Underwood, 1965; Postman, 1961; Underwood & Keppel, 1963). Thus, the theory which could claim considerable success in interpreting the results of the formal RI and PI paradigms was not supported and was, furthermore, left relatively impotent to account for the invariant amount of forgetting which did occur in the single-list experiments.

In any reevaluation of a theory that is reasonably successful in one context but not in another, several alternatives are always available. However, since the theory originated in the formal RI and PI situations, a reexamination of the effect of certain variables in these paradigms looked potentially fruitful. More particularly, our interest focused on the strength of interfering associations in the PI paradigm and the relationship between this strength and the amount of PI. It was assumed, in the unsuccessful tests of the theory, that linguistic habits of long standing will be extinguished by relatively few trials on a list where the associations conflict with the linguistic habits. This assumption may be invalid. Perhaps, as Mandler (1962) has argued, high levels of learning of associations result in these associations taking on a characteristic (structure) which prevents them from interfering with other associations. No study is available on the relationship between PI and degree of prior learning when the prior learning has been taken to a very high level and where the retention interval is relatively long.

In view of these considerations, the experiment was an attempt to simulate the effect on PI of associations which had experimentally been given characteristics presumed to be possessed by linguistic habits. Two such characteristics seemed potentially important. First, as noted, it must be assumed that linguistic habits are very strong and second, that linguistic habits have been developed by distributed practice over many years. Either or both of these characteristics might produce a reduction in PI despite the fact that formally the linguistic associations and those learned in the laboratory would appear to be in conflict (A-B, A-C). In the experiment, therefore, the degree of A-B learning has been varied from a low to an extremely high level. In one set of conditions the level of prescribed learning was given in a single session. In another set, the learning of A-B was distributed over 4 days. All *S*s learned a second A-C list which was recalled after 24 hours.

The thinking which prompted the experiment may now be summarized and the implications of several possible outcomes evaluated. It was noted that past experiments have been unsuccessful in demonstrating differential PI when associations learned in the laboratory are in differential conflict with extraexperimental linguistic associations. This is to say, the linguistic associations do not seem to "behave" in the same way as do associations established in the laboratory. It has been suggested that two quantitative differences between extraexperimental associations and laboratory associations may be involved in the discrepancy; namely, extraexperimental associations have been learned to a very high level, and they have been learned by distributed practice. The experiment, therefore, manipulated these two variables for A-B in a formal A-B, A-C paradigm in an attempt to have A-B simulate the extraexperimental associations. The influence of these two variables on the recall of A-C is examined, and, more particularly, it is asked if either of these variables is associated with a *decrease* in PI. Four relatively "pure" alternative outcomes of the experiment may be identified.

1. If degree of learning of A-B is the only critical factor, PI of A-C must decrease as degree of A-B learning increases, and the amount of decrease in PI must be independent of massing and distribution of A-B.

2. If the distribution of practice on A-B is the only critical factor, the reduction in PI resulting from distributed practice will be independent of the degree of learning of A-B.

3. If both variables are involved in a reduction in PI, maximum reduction should occur when A-B is learned to a high level by distributed practice if the relationships are monotonic. There may or may not be an interaction of the effects of the two variables.

4. Finally, neither variable may influence PI, in which case other sources must be examined for an understanding of why the interference theory has not been supported when applied to extraexperimental associations.

THE EXPERIMENT

Method

Lists. The two lists consisted of 12 paired associates in which the stimulus terms were three-letter words (medium frequency), and the response terms were two-syllable adjectives taken from Haagen's (1949) list. The lists formed an A-B, A-C paradigm, and the order of learning the lists was the same for all *S*s.

Procedure. The conditions may be thought of as representing two experiments, one in which the A-B learning was massed (MP), the other in which it was distributed (DP). There were four degrees of A-B learning in each experiment—12, 32, 48, or 80 trials. For the MP experiment the trials were all given in a single session and were immediately followed by the learning of A-C to one perfect trial. In the DP experiment, the A-B trials for each degree of learning were equally divided over 4 days. Thus, with 12 DP, *S* had 3 trials on each of 4 successive days; with 80 DP, *S* had 20 trials on each of 4 successive days. On the fourth day, following A-B learning, A-C was learned to one perfect trial.

Recall of A-C took place after 24 hours. *S* relearned for 5 trials or the number beyond this required to reach a perfect trial. Finally, a free-recall test was given. *S*s were supplied a sheet of paper on which the stimuli were listed, with two blank spaces under each, and were asked to supply the response terms from both lists and to identify (first or second) the list to which each response belonged. Unlimited time was given for the free recall.

Learning was by the anticipation method throughout, the rate being 2:2 seconds, with a 4-second intertrial interval. The words appeared in four different orders. The eight groups of college students were composed of 20 *S*s each; all *S*s had served in at least one verbal learning experiment prior to the present one. The assignment of *S*s to the conditions was random. It was recognized that the groups having high degrees of A-B learning, particularly the 48-MP and 80-MP groups, could become extremely bored with the task. Therefore special instructions were given which emphasized the importance of continuing to respond correctly on each repetition of the list. The 48-MP and 80-MP groups were promised an extra experimental point (toward the requirement imposed on all students in elementary psychology) for good cooperation. However, no *S* was informed as to the number of trials he would be given. A 5-minute rest was given the 80-MP *S*s after the first 40 trials on A-B.

RESULTS

A-B Learning

All eight groups had the first three trials on A-B under the same conditions; therefore, performance on these trials may be used to establish the comparability of the groups. The mean total-correct responses varied from 18.40–22.25, and all *F*s were less than 1.

It was of some importance that *S* continued to anticipate correctly in the overlearning trials if the differences in degree of learning are to be effective. The average number of times each item was given correctly was calculated. For the MP groups the values were 10.00, 29.48, 44.60, and 76.20 for 12, 32, 48, and 80 trials, respectively. Thus, out of a possible total of 80, the average pair was correctly anticipated 76.20 times for the MP-80 condition. The corresponding values for the DP groups were 9.68, 28.98, 45.24, and 77.32. For the higher degrees of learning, performance seems to have been maintained at a slightly higher level for the DP groups than for the MP groups. This may be shown in another way. The percentage of trials on which all items were correct was calculated after the trial on which all items had been first correctly anticipated. The values for MP 48 and MP 80 were 79.1% and 81.6%, and for the corresponding DP groups, 89.7% and 90.3%.

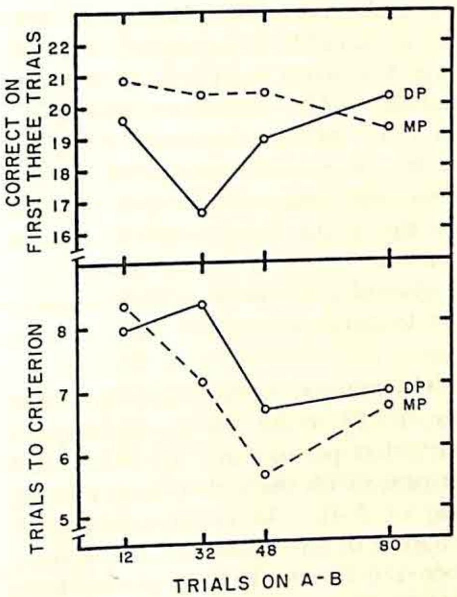

FIG. 1. A-C learning as a function of degree of A-B learning. (The upper panel gives the mean number of correct responses on the first three trials, the bottom panel, mean number of trials to one perfect trial.)

A-C Learning

Two measures of second-list learning (A-C) are plotted in Figure 1. The upper panel shows the mean number of correct responses on the first three trials, the lower panel, the mean number of trials to criterion. A statistical analysis of the data plotted in the upper panel shows no significant effect of either degree of learning or of MP-DP. The *F* for degree of learning is less than 1 (as is the interaction), that for MP-DP, 1.83, which is far from significant with $df = 1/152$. Analysis of the trials to criterion also fails to result in any variable being significant. The only F above 1 results from degree of learning (2.27), but with $df = 3/152$, an F of 2.67 is required for the 5% level. A trend analysis for unequal intervals (Gaito, 1965) still produced no significant component. It must be concluded that the wide variation in degree of A-B learning had no appreciable differential effect on A-C learning.

Certain considerations (e.g., Mandler, 1962) might lead to the expectation that performance on A-C would improve with the higher levels of learning of A-B. As noted, no statistical support of any change in A-C learning between 12 and 80 trials of A-B learning has been found. There is no control, such as A-B, C-D, to determine whether or not the transfer is positive or negative. However, if such a control were used, to produce improved performance in A-C with higher degrees of A-B learning would require that the C-D control show increasingly poorer performance as the number of trials on A-B increases. Available data (e.g., Postman, 1962) suggest that this would not be found.

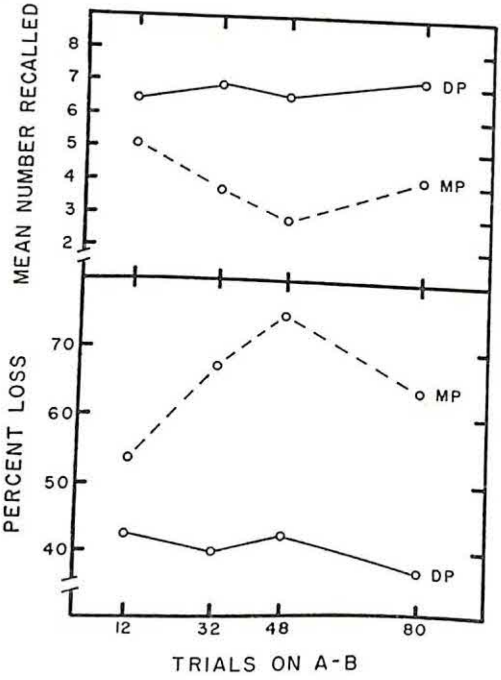

FIG. 2. Raw recall scores (upper panel) and adjusted percentage of loss scores (lower panel) as a function of degree of A-B learning.

It is known from another study (unpublished) that the two lists do not differ in difficulty. It was noted earlier that on the first three trials of A-B, performance varied between 18.40 and 22.25 correct responses. These are at best only slightly higher than the values seen in the upper panel of Figure 1. Across all conditions, the mean numbers of trials to reach one perfect on A-B were 7.06 for MP, 7.48 for DP. The corresponding values for A-C were 6.93 and 7.46. It should also be pointed out that if differences in learning to learn are associated with differences in degree of A-B learning, A-C performance should be increasingly facilitated by this factor alone. In short, no evidence is found of any marked transfer effects as a function of degree of A-B learning.

Finally, it may be noted that some intrusions of B responses occurred in learning A-C. Their numbers decreased as A-B learning increased, being 18, 15, 5, and 2 for MP; 11, 10, 6, and 1 for DP.

Recall

Two recall measures are plotted in Figure 2; the first to be considered will be the upper panel where the raw recall scores are plotted. An analysis of variance shows only one significant F, that for DP-MP (49.23). The F for degree of learning was 1.27, and a trend analysis does not change the conclusions. An analysis on the scores for the four MP groups alone produces an F of 6.30 for the quadratic trend with a value of 3.97 required for

TABLE 1

ERRORS ON THE RECALL TRIAL

Errors	Trials on A-B			
	12	32	48	80
Total MP groups	51	48	35	21
Intrusions from A-B	47	43	27	18
% intrusions	92.2	89.6	77.1	85.7
Total DP groups	28	15	15	10
Intrusions from A-B	17	5	9	3
% intrusions	60.7	33.3	60.0	30.0

the 5% level of significance. The linear ($F = 1.22$) and cubic ($F = .21$) trends were not significant.

There were small but insignificant differences in rate of learning A-C. The direction of the differences, however, suggested that recall differences might be underestimated somewhat, particularly for the MP groups. Therefore, a multiple-entry probability analysis was applied (Underwood, 1964) and percentage of loss scores derived for each *S*. The means of the loss scores are plotted in the lower panel of Figure 2. With the loss scores, the quadratic component of the degree-of-learning effect produces an F of 4.06 ($df = 1/152$), with an F of 3.91 required for the 5% level of significance. The quadratic interaction, however, falls considerably short of significance ($F = 2.80$). Nevertheless, any way of viewing the plots in Figure 2 must lead to the conclusion that recall is essentially uninfluenced by differences in degree of learning for the DP groups, but is influenced for the MP groups.

Plots of percentage of recall as a function of number of times A-C was given correctly in learning show that the DP groups have higher recall than the MP groups at all levels of response strength.

An analysis of trials to relearn A-C to one perfect recitation produced no significant source of variance, although the DP groups overall were somewhat superior to the MP groups (4.01 versus 4.61 trials).

Table 1 presents an analysis of errors on the recall trial. For both MP and DP, total errors decrease as degree of A-B learning increases. For the MP groups, the percentage of the errors which were intrusions varies from 77.1%–92.2%, with the percentages for the DP groups being appreciably lower.

Free Recall

For this measure the interest is in the recall of A-B following the learning and, after 24 hours, relearning of A-C. A lenient scoring system requires only that *S* produce the responses from A-B; a stringent scoring system requires that he also pair the response terms with the correct stimuli and correctly identify the pairs as to list. The results for both scoring systems are plotted in Figure 3. The DP groups, by either scoring method, show nearly perfect recall. By the stringent scor-

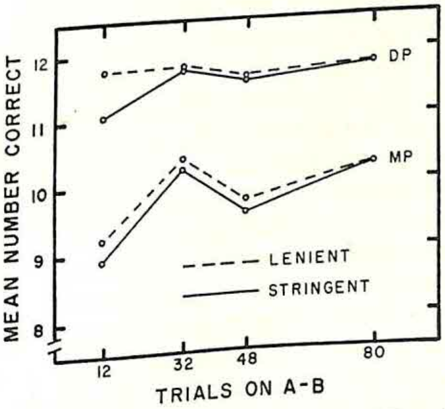

FIG. 3. Free recall of A-B for MP and DP groups as a function of degree of A-B learning.

ing system, only four, five, four, and four *S*s in the successive groups failed to show perfect recall. The near-perfect recall in the DP-12 conditions is particularly impressive when it is realized that two *S*s had failed to reach a perfect trial within 12 trials. The MP groups clearly show a loss, of approximately two items, but there is no difference as a function of degree of A-B learning ($F = 2.39$). There was not an appropriate control group to determine if the loss shown by the MP groups can be considered to represent significant extinction (significant RI). However, some estimates are possible. The DP groups on A-B learning provide evidence on 24-hour paced recall with no interpolated list after the initial blocks of trials, that is, after 3, 8, 12, and 20 trials on A-B. The recall scores for these four groups were 6.70, 8.50, 10.00, 10.55. As can be seen in Figure 3, free recall of A-B after 12 trials was approximately 9 items (versus 10 at the paced recall). With unpaced recall and no interpolated (A-C) learning it is suspected that the scores would probably be perfect with 32 or more trials. It is concluded that all MP groups show significant extinction effects.

IMPLICATIONS

Four relatively unambiguous possible outcomes of the experiment were identified in the introduction. As might be expected, the results do not match precisely any one of the four. Nevertheless, the second alternative, which stated that DP of A-B is the only critical factor involved in the reduction of PI, is reasonably accurate. The difference in the amount of PI for the DP and MP conditions is large and holds for all levels of A-B learning. If we accept the assumptions that (*a*) extraexperimental linguistic associations have been learned by DP, and (*b*) present DP conditions have adequately simulated this learning, previous failures to support the extension of interference theory to linguistic associations become more understandable. This finding does not absolve the theory; rather, it points to a deficiency, since the theory has no explicit provisions to account for the reduction in interference produced by distributed practice on the interfering list. This matter will be examined at a later point.

The effect of different degrees of learning of A-B on PI was somewhat ambiguous statistically, yet certain conclusions seem appropriate. There was no appreciable effect of different degrees of A-B learning under the DP schedule; recall following 80 A-B trials distributed over 4 days was essentially equivalent to the recall following 12 A-B trials distributed over 4 days. For the MP groups, however, degree of learning did influence recall. Recall decreased between 12 and 48 A-B trials, with a small increase apparent betwen 48 and 80 trials. Therefore, it is tentatively concluded that degree of A-B learning as such is not critically involved in the failure of extra-experimental associations to conform to expectations from interference theory.

Returning to the results for the DP groups, the question is whether or not there was any PI. That the outcome of the DP conditions could make the need for a control group (having no prior learning) so important was unforeseen. Unfortunately, when the need became apparent, there was no way to obtain the control *S*s with background experience in other verbal learning experiments that would duplicate the experience of *S*s in the DP groups. All *S*s in the DP groups had served in at least one previous experiment employing paired associates in which three-letter words were used, and most had served in two or more. The previous experiments in most instances had been conducted in the same room with the same memory drum by the same experimenters as the present one. Therefore, some proactive interference from these prior experiments must be expected, and this interference would have been present had a control condition been run. When a naïve control group is used, approximately 75% recall may be expected after 24 hours if the list is learned to one perfect recitation (Underwood, 1957). The question is, of course, how much this value would be reduced had a control been run consisting of *S*s with the same background of experience in verbal learning studies as those serving in the present conditions. Across all DP groups, recall was 56%, using 12 (number of items) as a base. We cannot be sure, of course, but we do not see how a control group could have recalled more than 65%, and perhaps 60% is more accurate. In any event, if there is PI shown in the recall of the DP groups, it is quite unlikely to be more than 10% in amount. Again the implications of the conclusion that there is little if any PI shown by the DP groups may be noted. If it is assumed that extraexperimental linguistic associations have been learned with distributed intervals far more extreme than those used here, it cannot be expected that such associations will interfere at recall with associations acquired in a few trials in the laboratory even though the associations learned in the laboratory are ostensibly in conflict (A-B, A-C) with the extraexperimental associations.

The critical question is why DP on the interfering list markedly reduces the interference produced by that list, whereas MP does not. No complete answer can be given at the present time, but there are certain contingencies in the data which may point to the critical mechanisms. The free-recall data for the DP groups indicated essentially no extinction of the A-B associations, even with the lowest degree of prior learning, whereas there was clear evidence for extinction in the MP groups. Yet, the A-C learning of the MP and DP groups did not differ as might be expected if extinction differed. That the DP groups showed no extinction is particularly impressive in view of the fact that there were two potential extinction periods (learning and relearning of A-C). There was no extinction in spite of the fact that intrusions from A-B were observed both in learning and relearning A-C, and intrusions have sometimes been associated theoretically with the process of extinction. The failure of extinction, therefore, is associated with a reduction in PI. It is as if potentially interfering associations learned by DP do not enter into an extinctive relationship in the same way that associations learned by MP do. Interference theory has assumed that spontaneous recovery (increasing strength) of the A-B associations is in part at least responsible for an increase in PI as the retention interval increases. Yet, the DP groups maintained the interfering associations at a high level of strength, but still there was little interference at recall. It is possible that the critical role of DP could still be related to extinction effects which reduce PI. Peripheral or weak extraexperimental associations tied to the A terms may be thoroughly extinguished by the DP schedule (Keppel, 1964), thereby leaving only the A-B association as a potential source of interference for A-C. Under such conditions, it may be possible to maintain the differentiation of the two response systems at recall. The use of other paradigms without common terms (e.g., A-B, C-D) would give evidence on the appropriateness of such an approach for revisions of the interference theory, revisions which seem necessary.

Although a direction has been suggested which may be taken by interference theory to accommodate the present findings, it is not intended that this be further pursued here. However, partly by way of summary and partly as a means of extending the implications of the present results, further comments will be made:

1. The interference theory of forgetting has been extended to include interference produced by associations learned outside the laboratory, particularly associations learned prior to laboratory learning. The basic purpose of this extension was to allow an accounting of the forgetting shown by a naïve *S* learning a single list in the laboratory. As noted in the introduction, tests of the theory have been rather uniformly negative in their attempts to show that materials differing in the degree to which they conflict with associations acquired outside the laboratory will be forgotten at different rates. The amount forgotten by a naïve *S* seems essentially independent of the nature of the material and the type of task and is sensitive only to the degree of learning. The present results strongly suggest that the source of interference cannot be linguistic habits of long standing, or, if they are, our methods of "bringing them into the laboratory" are quite inadequate. If we cannot identify specific sources of interference to account for the forgetting of the naïve *S*, we are left without an adequate accounting of the forgetting shown. When we attempt in the future to identify such interference, it appears that we must make it independent of the nature of the material learned in the laboratory; the relatively constant rate of forgetting of the naïve *S* across widely different materials and tasks does not provide an alternative. Until we are able to identify the source or sources of interference producing the constant forgetting in the naïve *S*, the decay theorist has some grounds for glee.

2. When an interfering list is distributed over days, the interference is markedly reduced. This cannot result from a difference in the strength of association as we normally think of strength, for clearly the 80-trial DP condition must have resulted in associations of greater strength than in the 12-trial DP condition, but still PI did not differ. When learned under MP, very strong associations still apparently interfered with the retention of A-C. The DP conditions not only resulted in fewer intrusions in learning and recall than did the MP conditions, but total errors at recall were quite different in frequency. One of the continuing problems of interference theory is "what to make" of errors, particularly intrusions. Are they to be taken as indicative of causal processes for forgetting, or are they merely the consequence of forgetting?

3. There have been recent discussions in the literature concerning the effect of overlearning on transfer and forgetting (Jung, 1965; Mandler, 1962, 1965). A significant effect on A-C learning as a consequence of differences in degree of A-B learning was not found. How can an A-B, C-D control provide data which, when the difference between C-D and A-C learning is examined, result in better A-C performance with high degrees of A-B learning than with low? The recall results for the DP groups gave no indication of differences as a function of degree of A-B learning. The only data which would tend to support a concept of structure in overlearning would be

the small reduction in PI with the MP groups between 48 and 80 A-B trials. Nevertheless, the large reduction in forgetting for the DP groups might be said to be in some way produced by a structure developed in A-B learning by DP and that this structure does not develop in learning by MP.

4. The present data show that widely distributed practice of A-B markedly reduces the forgetting of A-C. Comparable operations on A-C have also been shown to reduce markedly the forgetting of A-C (Keppel, 1964). If practice is widely distributed on both A-B and A-C, will forgetting be still further reduced, or is it the contrast (differentiation?) between practice schedules which makes distribution so effective (Houston & Reynolds, 1965)?

These problems, it seems, are some of the more important ones which must be handled, not only by an interference theory, but by any other type of approach. Some of the shortcomings in the details of interference theory have become gradually disconcerting over the past few years; the present study has merely verified certain of these shortcomings in a rather stark manner.

REFERENCES

EKSTRAND, B. R., & UNDERWOOD, B. J. Free learning and recall as a function of unit-sequence and letter-sequence interference. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 390–396.

GAITO, J. Unequal intervals and unequal *n* in trend analyses. *Psychological Bulletin*, 1965, **63**, 125–127.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association-value ratings of 400 pairs of common adjectives. *Journal of Psychology*, 1949, **30**, 185–200.

HOUSTON, J. P., & REYNOLDS, J. H. First-list retention as a function of list differentiation and second-list massed and distributed practice. *Journal of Experimental Psychology*, 1965, **69**, 387–392.

JUNG, J. Comments on Mandler's "From association to structure." *Psychological Review*, 1965, **72**, 318–322.

KEPPEL, G. Facilitation in short- and long-term retention of paired associates following distributed practice in learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 91–111.

MANDLER, G. From association to structure. *Psychological Review*, 1962, **69**, 415–427.

MANDLER, G. Subjects do think: A reply to Jung's comments. *Psychological Review*, 1965, **72**, 323–326.

POSTMAN, L. Extra-experimental interference and the retention of words. *Journal of Experimental Psychology*, 1961, **61**, 97–110.

POSTMAN, L. Transfer of training as a function of experimental paradigm and degree of first-list learning. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 109–118.

UNDERWOOD, B. J. Interference and forgetting. *Psychological Review*, 1957, **64**, 49–60.

UNDERWOOD, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 112–129.

UNDERWOOD, B. J., & KEPPEL, G. Retention as a function of degree of learning and letter-sequence interference. *Psychological Monographs*, 1963, **77**(4, Whole No. 567).

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychological Review*, 1960, **67**, 73–95.

(Received September 3, 1965)

Psychological Review
1966, Vol. 73, No. 6, 550–565

OPTICAL MOTIONS AND SPACE PERCEPTION:

AN EXTENSION OF GIBSON'S ANALYSIS[1]

JOHN C. HAY

Smith College

There is a perceptual tendency to see the 2-dimensional transformations of shadows as rigid movements in 3 dimensions. The significance of this illusion for veridical perception is investigated by a mathematical analysis of the correspondence between 3-dimensional object displacements and the optical motions they cause at the eye. The results show that optically equivalent object displacements share significant features, and that therefore the optical motions they produce are potentially informative about those features. Some of the objective features specified by optical motions are initial orientation, rotation, and the ratio between initial distance and translation distance. The analysis is based on coplanar sets of object points, and the fact that their displacements produce 2-dimensional projective transformations of the retinal image.

The sharply defined shadow of a stationary object, cast by a point source of light, may assume a variety of two- and three-dimensional appearances, like the black-and-white silhouettes of Gestalt psychology. But the transforming shadow of a moving object induces a more striking illusion: The shadow acquires a substantial appearance, and looks to be itself an object moving in depth. (Metzger, 1934; Wallach, 1953) The situation is illustrated in cross section in Figure 1A. It is as if the visual system, when presented with the transformation of the shadow, responded by resolving it into two components: a constant object structure and a change of position in space.

Gibson (1957) inferred from this phenomenon that the shadow preserves information about the object, not in its form, but in its transformation. "Information" is relative, of course, to a particular range of possibilities. Gibson proposed that the shadow transformation is informative with respect to the full range of *object displacements* (rigid motions); that is, rotation around, and translation along, each of the three dimensions of space. He noted, first of all, that the *optical motion* induced at the eye by the shadow is the same as would be induced by the object if the latter were viewed from a suitable station point. In developing his analysis, Gibson concentrated on the case of flat object surfaces. By watching the shadow transformations caused by displacing flat objects, he inferred the following correspondence between object displacements and optical motions:

Object Displacement		*Optical Motion*
Translation		
1. along line-of-sight.	→	Magnification.
2. along lateral (or vertical) axis.	→	Translation along lateral (or vertical) axis.

---

[1] This study is based on a doctoral thesis presented to Cornell University. Much of the work was supported by a United States Public Health Service Predoctoral Fellowship. J. E. Hochberg and J. J. Gibson provided invaluable counsel, and Louise S. Hay and E. Wong checked the original mathematical treatment.

Rotation	
3. around line-of-sight. →	Rotation around line-of-sight.
4. around lateral → (or vertical) axis.	Foreshortening along lateral (or vertical) axis.

Gibson hypothesized that the visual system is able to invert the above correspondence, and thereby identify object displacements from the optical motions registered at the eye. He found experimental evidence for this hypothesis by showing that subjects make certain kinds of judgments accurately by watching certain simplified shadow motions (Gibson & Gibson, 1957; Gibson, Gibson, Smith, & Flock, 1959; Schiff, Caviness, & Gibson, 1962).

The present study extends Gibson's analysis by a mathematical formulation of the correspondence between object displacements and optical motions. The results show that the above one-to-one correspondence is an oversimplification, and at the same time an underestimate of the potential informativeness of optical motions. There are eight relevant parameters of optical motion rather than the six in Gibson's list, and these parameters may specify the initial position of the displaced object, as well as the displacement.

PLAN OF THE ANALYSIS

The analysis will be organized around the question: How many different object displacements can produce one and the same optical motion? The size of these "optical equivalence" classes determines the discriminations that might be mediated by optical motions. The analysis will be similar to that provided by Ittleson (1960) for stationary objects. The emphasis will be, however, not on the size of the optical equivalence classes, but on their class characteristics, and thus on the informativeness rather than the ambiguity of optical stimuli.

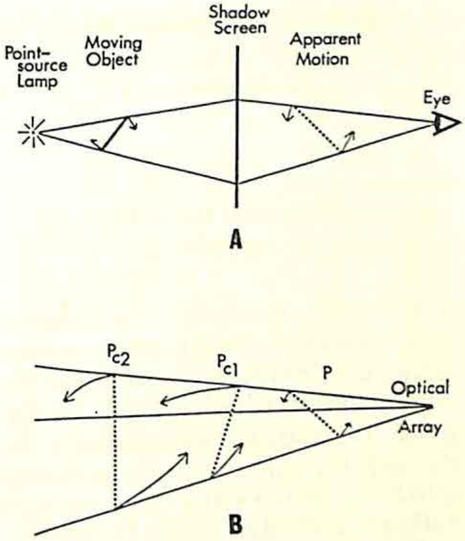

FIG. 1. A: Cross section of the moving shadow illusion. (The deforming shadow in the shadow screen produces the impression of a rotating object similar to the unseen shadow-casting object.)

B: Objects undergoing optically equivalent displacements in a two-dimensional cross section. (P is the true object causing the optical motion at the eye; P_{c1} and P_{c2} are two confusable objects undergoing optically equivalent displacements. Algebraic analysis is needed for the three-dimensional situation, where the class of confusable objects is much more restricted.)

No assumptions will be made about either the speed or the trajectory of the object displacements. Instead, only the geometrical relationship between the successive images that comprise an optical motion will be considered. The term *optical transformation* will be used to refer to the relationship between two successive stages of an optical motion.

Figure 1B shows the case of several different object displacements producing the same transformation of the *optical array* (the sheaf of light

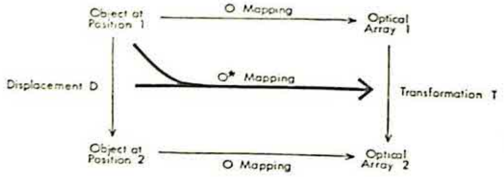

FIG. 2. Diagram of the mathematical analysis.

rays entering the eye). The diagram is a two-dimensional cross section of the real situation. P stands for the "true" object of Figure 1A, undergoing the "true" displacement D. P_{c1} and P_{c2} stand for two *confusable objects*, in the sense that they can each undergo a displacement, D_{c1} or D_{c2}, which causes the same optical transformation as P and D. The situation is comparable to the familiar case of different stationary objects being optically equivalent with respect to the eye (e.g., Ittleson, 1960, p. 67). If the two-dimensional cross section of Figure 1B were an adequate representation of the three-dimensional situation, our question would already be answered: Many different object displacements are optically equivallent, in cross section, to one and the same optical transformation.

In the real three-dimensional situation, the optical equivalence classes will be shown to be much more restricted than suggested by the cross section of Figure 1B. For this analysis, algebraic techniques are more convenient than geometrical constructions. Matrix algebra at the level developed by Birkhoff and MacLane (1965) will be used. To facilitate understanding, a verbal summary of the algebraic argument follows.

Figure 2 shows the analysis in abstract form. To begin with, the *O Mapping* is defined, which maps objects into optical arrays at the eye's station point. This O Mapping does not preserve three-dimensional structure, since the optical array (and the image that it defines on the retina) is essentially two-dimensional.

The central problem of the analysis concerns the O* Mapping in Figure 2. This is the mapping of object displacements, D, into optical transformations, T. First it is shown that the O* Mapping is determined jointly by D and by the initial position of the object; thus the diagram in Figure 2 shows two "inputs" for this mapping. Second, it is shown that if the object is a flat surface, the O* Mapping causes a specific geometrical transformation of the optical array: a two-dimensional *projective transformation.* Projective transformations have definite properties of their own, independent of the optical array to which they are applied, just as one may specify an object displacement without specifying the object. Finally, it is shown that for a multifaceted object, the O* Mapping causes distinct projective transformations for each flat surface of the object, when the object as a whole is displaced. That is, the complicated optical transformation caused by displacing a solid object can be analyzed into a family of distinct projective transformations, $PT_1 \ldots PT_n$, one PT corresponding to each face or coplanar set of points in the object. This last finding, based on what will be termed the Parallax Theorem, suggests that further analysis concentrate on projective transformations of the optical array.

The question with which we began now becomes: Is there a narrowly limited equivalence class of surfaces and displacements for every projective optical transformation? The affirmative answer to this question is contained in the Distance and Orientation Theorems.

The final problem dealt with concerns the properties of projective optical transformations and how they might be registered by the eye. For this problem, the optical array is converted into an image on a flat plane tangent to the retina at the fovea. In terms of this image plane, eight parameters of projective transformation are specified, and they are shown to bear a complicated relationship to the parameters of object surface and displacement. It is further indicated how eye movements, by fixating or tracking certain parts of the transforming array, may simplify the image movement actually impinging on the retina and thereby facilitate its registration.

THE MAPPING OF OBJECT DISPLACEMENTS INTO OPTICAL TRANSFORMATIONS

The following symbolic conventions will be used:

a, b, $\cdots$	= scalars
$a, b, \cdots$	= 3-unit row matrices; for example, $(a_1\ b_1\ c_1)$
$A, B, \cdots$	= 3×3 square matrices
a', A', etc.	= transposed matrices
AB, aB, ab', etc.	= matrix products, multiplying the rows of the left-hand matrix by the columns of the right-hand matrix.
$[\mathrm{n}a], [\mathrm{n}A]$, etc.	= the set of all scalar multiples of a, A, etc.

The following correspondences between matrices and geometrical concepts will be employed:

$a, b, \cdots$: vectors (or points in 3-space) relative to a set of coordinate axes.

$aa' = a_1^2 + a_2^2 + a_3^2$: squared length of vector $a = |a|^2$

$ab' = a_1b_1 + a_2b_2 + a_3b_3$: inner product of vectors a and $b = |a|\,|b| \cos \angle ab$.

$ap' = 1$: a plane P, whose inverse intercepts comprise p.

$aCC'a' = 0$: a quadratic surface in space, orthogonally equivalent to the points satisfying the equation $c_1a_1^2 + c_2a_2^2 + c_3a_3^2 = 0$, where c_1, c_2, and c_3 are the characteristic roots of CC'.

Figure 1B can now be interpreted in three-dimensional space by setting normal orthogonal base-vectors on the eye's nodal point: a *lateral* or *X-axis* vector perpendicular to the diagram; a *vertical* or *Y-axis* vector running vertically in the diagram; and a *line-of-sight* or *Z-axis* vector running horizontally. Each object point in space now corresponds to a vector $e = (x_e y_e z_e)$.

The optical array will be represented in terms of a single pencil of light rays converging at the eye's nodal point, as in Figure 1B. Since the nodal point is the origin of the vector space, each ray is a one-dimensional subspace; that is, the set $[\mathrm{n}e]$ where n ranges over the scalars.

With these definitions, the O Mapping in Figure 2 of object points into optical rays becomes:

$$\mathrm{O}(e) = [\mathrm{n}e] \qquad [1]$$

This constitutes a singular perspective mapping of three-dimensional space into the two-dimensional space of the optical array (cf. Klein, 1908). It is clearly a many-to-one mapping: Any two object points e_1 and e_2 have the same mapping so long as $e_2 = \mathrm{n}e_1$ for some scalar n. The mapping has one important restriction: Object points behind the eye do not reflect light into it. Therefore there is the following qualification on Equation 1:

$$\text{If } O(e) \text{ exists, then } z_e \geq 0$$

O is not, however, the needed mapping of object displacements into optical transformations. This mapping is O* of Figure 2, and is derived indirectly as follows:

The displacement D in Figure 2 is, by definition, a rigid motion, a change of position without change of size or shape. Every such transformation is a combination of a rotation and a translation, and can be represented as follows (Birkhoff & MacLane, 1965, p. 243):

$$D(e) = eR + t \qquad [2]$$

where R is orthogonal, its elements being the direction cosines of the rotation, and t is a row matrix, its elements being the X, Y, and Z components of the translation.

Having thus determined the O Mapping and the Displacement D in Figure 2, it is possible to derive the algebraic expression for the Transformation T of the optical array. The initial optical array produced by the object with points e is given by the O Mapping; that is, the rays $[ne]$. The transformed Optical Array 2 in Figure 2 is given by applying D and O in sequence to the object points e:

$$\begin{aligned} T([ne]) &= OD(e) \\ &= O(eR + t) = [neR + nt] \qquad [3] \end{aligned}$$

This equation, representing the general case of a three-dimensional displacement of a three-dimensional object, is difficult to analyze. Far from being a one-to-one transformation, T is "many-to-many": $T([ne])$ is not uniquely determined by $[ne]$, nor vice versa.

If attention is limited to coplanar points of the object, however, the optical transformation T can be analyzed into simpler components. It will be shown, first, that the optical transformations caused by displacing coplanar points are of a special category, namely the group of projective transformations. It will then be shown that the general optical transformation T can, at least in many cases, be expressed as a set of projective transformations, $PT_1 \cdots PT_n$.

All coplanar points satisfy an equation of the following form:

$$ep' = 1 \qquad [4]$$

where the elements of p are the inverse intercepts of the plane P on the X, Y, and Z axes. This means that for the flat surface P the displacement (Equation 2) is equivalent to a linear transformation of space represented as follows:

$$D(e) = eA \qquad [5]$$

for all e such that $ep' = 1$ and where $A = R + p't$. (Note: $p't$ is a 3×3 matrix of rank 1.) Combining Equations 2, 3, and 5 gives the new optical transformation PT, which is produced by displacing coplanar object points:

$$PT[ne] = [neA] \qquad [6]$$

Equation 6 satisfies the definition of a projective transformation (Birkhoff & MacLane, 1965, p. 277). Such transformations are uniquely determined by only four rays (so long as no three of them lie in the same plane through the origin) and their transforms.

The mapping O* of object displacements into optical transformations can now be defined for coplanar object points, in exact parallel to the mapping O of object points into the optical array. For this purpose, it is necessary to note that two matrices A_1 and A_2 will determine the same projective transformation in Equation 6 if and only if $A_2 = \mathrm{m}A_1$ for some scalar m (Birkhoff & MacLane, 1965, p. 277). Thus a projective transformation is represented by a set of matrices $[\mathrm{m}A]$, just as an optical ray is represented by a set of matrices $[\mathrm{n}e]$. In terms of the matrices which represent object displacement (R and t) and the object plane (p):

$$\begin{aligned} \mathrm{O}^*(R,t,p) &= [\mathrm{m}R + \mathrm{m}p't] \\ &= [\mathrm{m}A] \end{aligned} \qquad [7]$$

The question to be raised in the next section is whether O* is a one-to-one mapping, unlike O. It remains to be determined here, however, whether the foregoing formulation of O* in terms of coplanar object points can serve to analyze the general optical transformation of Equation 3. Consider the case of a multifaceted object; that is, one with several plane surfaces. Displacing such an object will cause a set of projective transformations of the optical array. Will there be a distinct projective transformation for each distinct face of the object? Furthermore, could nonplane surfaces of the object also induce projective transformations of the optical array? These questions can be answered as corollaries to what may be called the Parallax Theorem.

Parallax Theorem

If two object points map into the same optical ray, they will no longer do so after they have both been subjected to a displacement, except under specified circumstances.

Proof. If two points e_1 and e_2 have the same optical array mapping, then $e_2 = \mathrm{n}e_1$ for some scalar n (Equation 1). Now suppose that their transforms after displacement were to still map into a common optical ray. This implies:

$$\begin{aligned} e_1R + t &= \mathrm{m}(e_2R + t) \\ &= \mathrm{m}(\mathrm{n}e_1R + t) \end{aligned}$$

This vector equation can be rewritten as follows:

$$(1 - \mathrm{mn})e_1R = (\mathrm{m} - 1)t \qquad [8]$$

Equation 8 can be true only under two circumstances: (*a*) if e_1R lies in the same direction from the eye as t, or (*b*) if t is zero.

Corollary 1. When a nonplane set of object points is displaced, it does not cause a projective transformation of the optical array, except under specified circumstances.

Proof. Assume a set of five coplanar object points: $e_1 \cdots e_5$. When subjected to a displacement D, they induce a projective transformation PT on the optical array, where PT $= [\mathrm{n}R + \mathrm{n}p't]$ (Equation 7). Now assume that e_5 is replaced by another object point e_6 which has the same O mapping as e_5, but does not lie in the plane of $e_1 \cdots e_4$. If e_6 is to induce the same projective transformation as $e_1 \cdots e_4$, then De_6 must have the same O mapping as De_5. But this contradicts the above theorem, with the exceptions noted for that theorem. A further exception occurs for the corollary when there are only four object points: Since any four rays and their transforms determine a projective transformation, those rays need not originate in coplanar object points.

Corollary 2. When n distinct object planes are subject to a common displacement, they induce n distinct projective transformations of the optical array, except under specified circumstances.

Proof. Corollary 1 showed that nonplanar points cannot, in general, cause a single projective transforma-

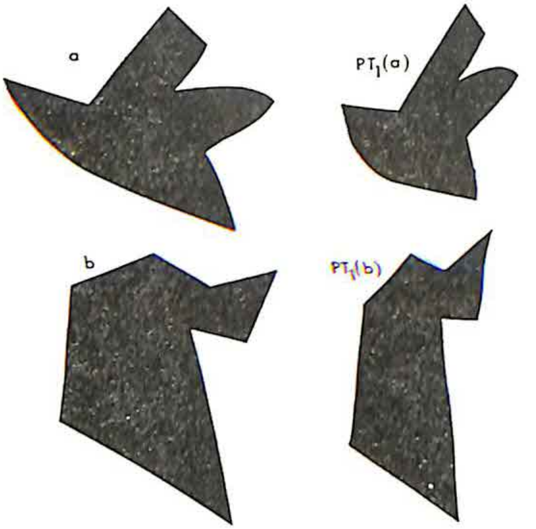

FIG. 3. Two forms (a and b) undergoing the same projective transformation, PT_1. (PT_1 corresponds to a flat surface rotating from 5° out of the frontal plane to 55° from frontal, around an axis tilted slightly clockwise from vertical.)

tion of the optical array. But Equation 6 shows that every set of coplanar object points induces a projective transformation of the array. Hence every distinct set of coplanar points induces a distinct projective transformation of the array. The exceptions to this are the same as to Corollary 1.

Discussion

The analysis has shown that, just as objects are mapped into optical arrays, displacements of objects (so long as the objects are flat) are mapped into projective transformations of the optical array. Whatever the form or texture of the flat surface, so long as it undergoes a specific displacement from a specific position, it will produce the same projective transformation in the optical array. (See Figure 3.)

All the optical rays originating in the flat object surface will "move together," not in the sense that their separations remain constant, but in the sense that if three rays lie in the same plane they will remain coplanar. That is, in terms of the retinal image formed by the rays, straight image lines will remain straight, while their lengths and orientations vary. If the visual system is able to respond selectively to optical motions that preserve straight image lines, then it will have the basic capacity for discriminating spatial displacements of flat surfaces.

Nonflat object surfaces cannot induce projective transformations of the array, except under very special circumstances (Parallax Theorem, Corollary 2). If the object is multifaceted, each of its surfaces will induce a separate projective transformation of the array (Corollary 2). In this fashion, the complicated optical transformation, T, caused by displacements of three-dimensional objects can be analyzed into a family, $PT_1 \cdots PT_n$, where n is the number of flat surfaces in the object. (See Figure 4.) Objects with curved surfaces may also be so treated, in terms of planes which intersect them at many points, or in terms of tangent planes which approximate their surfaces.

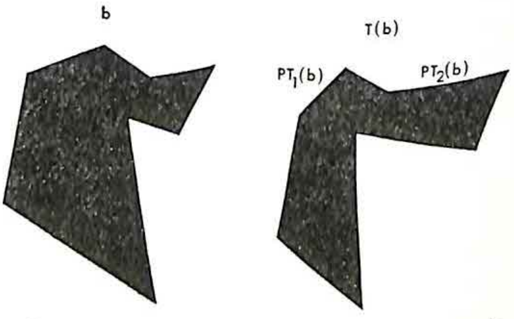

FIG. 4. Left and right halves of Form b in Figure 3 undergoing two different projective transformations, PT_1 and PT_2. (The effect corresponds to a 50° rotation of an object surface with a 70° bend in it.)

THE SPECIFICITY OF OPTICAL TRANSFORMATIONS TO OBJECT DISPLACEMENTS

The algebraic model provides a means for answering the question with which we began: How many different object displacements are optically equivalent to one and the same transformation of the optical array? The question can now be put in terms of projective transformations, and their matrix representation. In brief, the question now consists of asking in how many different ways a 3×3 matrix A can be expressed as the sum of an orthogonal matrix R and a rank 1 matrix $p't$ (Equation 5). The result of the analysis will show that, in the general case, a given projective transformation can be caused by two different object displacements, each applied to a different initial surface orientation. Intuitively, the argument depends on the fact that object displacements preserve distance between object points; that this requirement leads to a quadratic locus; and that for a quadratic locus to include a flat surface means that it must be, in the general case, two planes.

Consider the question as put schematically in Figure 1B: Given a transformation of the optical array caused by the displacement of the flat surface P, how many other surfaces P_c could be displaced in an optically equivalent fashion? Each confusable surface P_c is specified by a column matrix p'_c, just as P was specified by the matrix p (Equation 4). The displacement of P_c is specified by an orthogonal matrix R_c and a row matrix t_c (cf. Equation 2). The displacement of P_c is therefore equivalent to the linear transformation $A_c = R_c + p'_c t_c$ (cf. Equation 5). For notational convenience, we shall continue to use p, R, t and A without subscripts as the corresponding values for the true surface P.

In order for the confusable surface P_c to cause the same optical transformation as the true surface P, Equation 6 sets the following condition:

$$A_c = \mathrm{m}A \quad \text{or} \quad R_c + p'_c t_c = \mathrm{m}(R + p't) \qquad [9]$$

There is a restriction on this equation, deriving from the restriction on the O mapping (Equation 1): The scalar multiple m must be positive, or else A_c would transform object points to the region behind the eye.

Two theorems will be proved about Equation 9, and taken together they provide the answer to the above question.

Distance Theorem

If the confusable surface P_c is parallel to the true surface P, then it will induce the same optical transformation as P if and only if $R_c = R$ and $t_c = \mathrm{h}t$, where h is the ratio between the distances of P_c and P from the eye.

Proof. Let P be determined, as above, by $ep' = 1$, and P_c by $ep'_c = 1$. If the planes of P and P_c are parallel, and the points in P_c h times further from the eye, then $p_c = (1/\mathrm{h})p$. Consider now the separations between object points either in P or in P_c: All such separations are vectors of the form $d = e_1 - e_2$, where e_1 and e_2 both lie in P, or both in P_c. Due to the distributivity of matrix multiplication, $dp' = e_1p' - e_2p'$, from which we infer the following:

$$dp' = 0 \qquad [10a]$$

$$d_cp'_c = 0 = d_c(1/\mathrm{h})p' = d_cp' \qquad [10b]$$

In short, d and d_c both lie in the plane through the origin designated by $dp' = 0$. The proof now depends on the fact that when P or P_c is displaced, that is, moved without change in size or shape, the length of the separation vectors d in this plane must remain unchanged:

$$(dA_c)(dA_c)' = (dA)(dA)' = dd' \qquad [11]$$

From Equation 9 and the distributivity of scalar multiplication it follows that $A_c = A$. To go from this to $R_c = R$ we need to note that $dA_c = dR_c$ and $dA = dR$ (combine Equations 9 and 10). Therefore $dR_c = dR$ for all vectors d lying in the plane $dp' = 0$. But the rotation of a plane determines a unique rotation of all space, since rotation preserves the length and sign of vector cross-products. Therefore $R_c = R$. From this it follows that $p'_c t_c = p't$, and $t_c = ht$ trivially.

Corollary 1. Given any component R, p, or t, the other components of a matrix A are determined, at least to within a scalar multiple.

Proof. If p is determined, the theorem proves the corollary. If R is determined, clearly $p't$ is determined by subtraction from A. The columns of $p't$ are scalar multiples of p'; the rows are scalar multiples of t. If t is determined, finally, the same algebraic argument as in the above theorem can be made; that is, for all b' such that $tb' = 0$, $Ab' = Rb'$, and R is determined as above.

Interpretation of the Distance Theorem

This theorem shows that the optical transformation is in no way specific to the particular distance of the surface P or to the length of the translatory component of D, but that the optical transformation is specific to the ratio between those two quantities. A family of flat surfaces differing only in their distance from the eye have a common *surface orientation*, and this property is specified by the column matrix p' (or any scalar multiple thereof) in the equation $dp' = 0$. (Contrast this with the particular surface specified by p' in Equation 4, $ep' = 1$.)

It remains to be shown whether a projective optical transformation is specific to a particular surface orientation. To investigate this, the following lemma will be needed:

Lemma 1. If $A = R + p't$ for some matrices R, p', and t, then only A in the set $[\mathrm{m}A]$ is equivalent to a displacement of an object plane.

Proof. The proof follows from a closer examination of Equation 11, which can be rewritten as follows:

$$d(\mathrm{m}^2AA' - I)d' = 0 \qquad [12]$$

where I is the identity matrix. For every value of m, m^2AA' is a positive definite matrix (Birkhoff & MacLane, 1965, p. 256), and hence m^2AA' has three positive characteristic roots, and $(\mathrm{m}^2AA' - 1)$ has three characteristic roots. Thus Equation 12 is satisfied by a quadratic surface in space, and without loss of generality may be written (relative to a suitable orthogonal basis) as follows:

$$(\mathrm{m}^2\mathrm{u} - 1)\mathrm{X}^2 + (\mathrm{m}^2\mathrm{v} - 1)\mathrm{Y}^2 + (\mathrm{m}^2\mathrm{w} - 1)\mathrm{Z}^2 = 0 \qquad [13]$$

where u, v, and w are the (positive) roots of AA'. But the vectors d must also satisfy Equation 10; that is, they lie in a plane. Equation 13 is satisfied by a plane only if at least one of the coefficients is zero, and the other coefficients (if more than one) are of opposite algebraic sign. Thus if v (or u, or w) is the middle-valued root of AA', $\mathrm{m}^2\mathrm{v} = 1$. This can be true for only one value of m, since v is fixed, and $\mathrm{m} \geq 0$ (restriction on Equation 9). By the hypothesis of the lemma, a plane satisfies Equation 13 *if* $\mathrm{m} = 1$; hence this is so *only if* $\mathrm{m} = 1$. Q.E.D.

Orientation Theorem

In the general case, a projective transformation of the optical array is specific to two alternative surface orientations, and two corresponding displacements. The exceptions to this general case will be specified.

Proof. A projective transformation corresponds to a set of matrices $[mA]$ (Equation 7). But by the foregoing lemma, only A, as defined in the lemma, corresponds to an object displacement. Hence the present theorem depends on showing that A can be decomposed in only two ways into the sum of an orthogonal matrix and a rank 1 matrix $p't$. To show this, we note that Equation 13 sets limits on p'. No plane satisfies Equation 13 unless one of its coefficients is zero. If v (or u, or w) is the middle-valued root of AA', as in Lemma 1, the equation is simplified:

$$(u-1)x^2+(w-1)z^2=0 \qquad [14]$$

The following cases arise:

Case 1: The general case is when $u>1$ and $w<1$, or vice versa. In this case, Equation 14 is satisfied by two planes passing through the origin, one of which must be the true plane $dp' = 0$, and the other is the confusable plane $dp'_c = 0$. By Corollary 1 of the Distance Theorem, p_c specifies the optically equivalent displacement, D_c. The confusable surface p_c bears the following relationship to p, as can be seen by writing Equation 12 in scalar form, letting $dp' \neq 0$ and canceling that term:

$$p_c = |t|^2 p + 2tR' \qquad [15]$$

where $|t|$ is the length of t.

Case 2: If u or w is equal to one, then only one plane, $dp' = 0$, will satisfy Equation 14, and there is no confusable surface orientation. From Equation 15 it may be deduced that the absence of a confusable surface orientation depends on the following special relationship between p and the displacement:

$$\frac{m - |t|^2}{2}\, p = tR' \qquad [16]$$

for some scalar $m \neq 0$.

Case 3: If both u and w are equal to one, then all points in space will satisfy Equation 14, and the resulting projective transformation will bear no unique relationship to an object displacement. From Equation 15 it can be deduced that this depends on the following alternative conditions:

$$p=O \quad \text{or} \quad t=O \quad \text{or} \quad m=0 \qquad [17]$$

where O is that vector with all elements zero.

Corollary 1. If a multifaceted object is displaced, thereby inducing a set of simultaneous projective transformations of the optical array, that set of projective transformations will specify the object displacement, and the orientations of the object faces.

Proof. Under the hypothesis of this corollary, each of the projective transformations will have A matrices of the form:

$$A_n = R + p'_n t$$

Any two of these matrices have a difference of the following form:

$$A_1 - A_2 = (p'_1 - p'_2)t, \text{ etc.}$$

That is, two of the projective transformations specify t. By Corollary 1 of the Distance Theorem, this specifies R, and also p'_1 and p'_2, etc.

Corollary 2. If a single set of coplanar object points is present, and is subjected to a sequence of displacements, as in the usual case of a continuous object motion, there will be a confusable surface orientation only under specified circumstances.

Proof. Under the hypothesis of this corollary, the optical array is subjected to a sequence A_1, A_2, $\cdots$ of projective transformations. A_1 and A_2 apply different displacements to the same true surface P. For their confusable surfaces p_{c1} and p_{c2} to be identical, the following is required by Equation 15:

$$|t_1|^2 p + 2t_1 R'_1 = |t_2|^2 p + 2t_2 R'_2 \quad [18]$$

This can be rewritten as a condition on the rotation and translation components of A_2:

$$t_2 R'_2 = 1/2(gp + 2t_1 R'_1) \quad [19]$$

where g is an arbitrary scalar. This clearly requires that the product $t_2R'_2$ lie in the plane of p and $t_1R'_1$. Except in that special circumstance, A_1 and A_2 will have no confusable surface in common.

Interpretation of the Orientation Theorem

This theorem shows that, in the general case, the optical transformation caused by displacing an object plane is specific to two alternative surface orientations, and two corresponding displacements. Thus, in the general case, there is one confusable surface orientation, P_c.

A one-to-one correspondence is derived from this two-to-one correspondence when the object undergoing displacement includes two or more sets of coplanar points, or when a single set of coplanar points undergoes a sequence of displacements.

Discussion

When a flat object surface moves, it causes a continuous motion of the optical array. This optical motion may be termed a *projective optical motion*, since it will be composed of a continuous sequence of projective transformations. Any one of the transformations is sufficient to specify two alternative object surfaces, and two corresponding displacements (Orientation Theorem), but not absolute distance (Distance Theorem). Any two of the transformations will suffice, in the general case, to specify a unique surface orientation for the initial position, and consequently a unique sequence of displacements (Corollary 2 to the Orientation Theorem). Thus the projective optical motion caused by the moving flat surface contains a great deal of redundant information concerning the object surface and its motion. This makes it likely that even a low degree of precision in registering projective transformations would make possible, over a sufficient exposure period, an accurate identification of object surface and motion.

In the case of an object with several flat surfaces, the family of projective transformations produced by any one object displacement will suffice, in the general case, to specify that displacement uniquely, and consequently the set of object surfaces (Corollary 1 to the Orientation Theorem). Assuming that the visual system could register concurrent projective transformations, this means that multifaceted objects produce an even greater degree of redundant information during optical motions, so that perception of such objects should be more accurate than the perception of single flat surfaces.

The case of objects with curved surfaces is not so clear. Since there is a continuous range of planes which intersect such surfaces, there must be a continuous range of projective transformations induced by such surfaces. It seems likely that registration of such transformations would be difficult, and consequently a relatively low accuracy would be predicted for the perception, by means of optical motions, of curved surfaces and their movements.

PARAMETERS OF THE PROJECTIVE OPTICAL TRANSFORMATIONS

What are the properties of projective optical motions? That is, what would be involved in their registration by a visual system, and their "decoding" back into three-dimensional objects and displacements?

The properties are most conveniently specified not in terms of the optical array, but in terms of the image which the array produces on the retina. The image may be regarded, geometrically, as the intersection of the rays comprising the array with the two-dimensional surface of the retina. If only a small region (e.g., the fovea) is considered, it may be approximated by a tangent plane. This *image plane*, containing image vectors i, is specified by $iv' = 1$ for some fixed matrix v'. If the fovea is taken to lie at unit distance from the nodal point of the eye, then the "foveal image plane" is $z = -1$. (When the eye turns, or when a different region of the retina is considered, it is equivalent to rotating the image plane.)

In terms of the foveal image plane, Equation 6 can be written out in scalar form as follows:

$$\begin{aligned} PT(x) &= \frac{ax + by - c}{gx + hy - m} \\ PT(y) &= \frac{dx + ey - f}{gx + hy - m} \\ PT(z) &= z = -1 \end{aligned} \qquad [20]$$

The number of coefficients in these scalar equations is nine; however, only eight of them are independent parameters, since numerator and denominator of both fractions can be multiplied by a scalar without altering their function values. Hence it may be concluded that there are eight parameters of projective image transformation induced by displacing an object plane. This result may be compared with Gibson's identification of only six parameters of image transformation (1957, p. 290). The discrepancy arises because Gibson was only looking for correspondents of six parameters of displacement, and neglecting the two parameters of surface orientation at the start of the displacement.

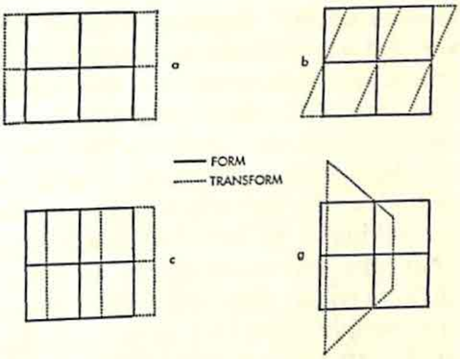

FIG. 5. The four horizontal parameters of projective transformation in the image plane. (a is stretching, b is shearing, c is translation, and g is foreshortening. The four vertical parameters can be seen by rotating the Figure by 90 degrees.)

The eight parameters can be sorted into four types, each applying to both lateral (X-axis) and vertical (Y-axis) dimensions of the image. They are illustrated in Figure 5.

Image translation (c and f parameters; Figure 5c).
Image stretching (a and e parameters; Figure 5a).
Image shearing (b and d parameters; Figure 5b).
Image foreshortening (g and h parameters; Figure 5g).

Image stretching bears special comment. Equal vertical and lateral stretching, that is, magnification, is produced by the parameter m in

Equation 20. Separate lateral and vertical stretchings are produced by the parameters a and e respectively. However, it is clear that these three coefficients only result in two independent image changes, which we will describe as lateral and vertical stretching. It is also noteworthy that the transformations of lateral and vertical shearing, when equal, and supplemented by suitable stretchings, produce rotation of the image. Our listing of the parameters of image transformation is thereby coordinated with Gibson's (1957) list: He identified only the two factors, magnification and rotation, and did not distinguish their horizontal and vertical components of stretching and shearing.

A visual system capable of detecting these four parameters, both laterally and vertically, would have picked up the information contained in projective optical transformations. However, it may be asked whether these parameters bear a simple and direct relation to the parameters of object displacement, as proposed in Gibson's (1957) analysis. It can readily be shown that they do not.

Equation 5, when written out in scalar form, shows how the optical transformation parameters are related to those of the object. Rather than writing out the cumbersome, but elementary, algebra, a qualitative account of the relationship will be given here. For this purpose, the following parameters will be used to describe the object and its displacement:

Initial surface orientation: measured relative to the frontal plane. *Lateral slant* is measured by the slope of the surface's intersection with the XZ plane; *vertical slant* by the slope of the surface's intersection with the YZ plane.

Initial surface distance: measured by the intercept of the surface with the line-of-sight (Z-axis).

Rotation axis: measured in relation to the three coordinate axes. The *vertical rotation component* is the projection of the rotation axis on the vertical coordinate axis, and similarly for the *lateral* and *line-of-sight rotation axes.* (The angular magnitude of the rotation will not be taken account of in this qualitative description.)

Translation axis: measured by its lateral, vertical, and line-of-sight components. (It can easily be shown that any object rotation around an axis which does not pass through the eye involves a translation.)

In terms of these object parameters, Equation 6 can be written out in the following way:

Lateral Image Translation =

$$\text{Vertical Rotation Component} + \frac{\text{Lateral Translation}}{\text{Surface Distance}}$$

Lateral Image Stretching =

$$\begin{array}{c}\text{Line-of-sight or}\\ \text{Vertical Rotation Component}\end{array} + \frac{\text{(Lat. Trans.) (Lat. Slant)}}{\text{Surface Distance}}$$

Lateral Image Shearing =

$$\text{Line-of-sight Rotation Component} + \frac{\text{(Lat. Trans.) (Vert. Slant)}}{\text{Surface Distance}}$$

Lateral Image Foreshortening =

$$\text{Vertical Rotation Component} + \frac{\text{(L-of-s Trans.) (Vert. Slant)}}{\text{Surface Distance}}$$

For vertical image transformations, interchange the terms "vertical" and "lateral" wherever they appear in the above equations. For simplicity, the case of equal lateral and vertical stretching, or magnification, can be listed separately:

$$\text{Image Magnification} = \text{Lateral or Vertical Rotation Component} + \frac{\text{Line-of-sight Translation}}{\text{Surface Distance}}$$

From these equations it is clear that there is far from a direct one-to-one correspondence between the parameters of image transformation and those of object displacement. Indeed, the relationship may be summarized by saying that each type of image transformation is influenced by all four types of object parameter, and by all three dimensions of space. It follows from this that detection of the parameters of an image transformation would need to be followed by an interaction between them, in order for the object characteristic to be identified.

Effects of Eye Movement

Since the eyes are free to move, it is possible that by suitable rotations the parameters of an image transformation may be simplified on the retina, either in the sense that some of them are made zero, or in the sense that they are made to bear a more direct relationship to the object motion than is true of the general case illustrated above. For example, reflex pursuit movements of the eye will tend to hold foveal image points stationary on the retina, and this is equivalent to zeroing the translation parameters of the optical motion. A detailed analysis of the possibilities will not be given here, but two will be sketched:

Choice of fixation point. If the eye is turned so as to adopt a suitable posture relative to the axes of the object and its displacement, the parameters of the image transformation may be simplified. For example, if there is no object rotation, and the eye is turned so that the line-of-sight lies along the translation axis, stretching and foreshortening will be the only image transformations. This is essentially the case considered by Gibson, Olum, and Rosenblatt (1955). As a further example, if the rotation axis is fixed in space, and the eye turned so that (*a*) the vertical axis is parallel to the rotation axis and (*b*) the line-of-sight axis is pointed in the direction of the rotation axis, there will be no vertical shear or translation. Algebraically, such changes of fixation can be represented as orthogonal changes of the coordinate system, changing the transformation matrix A into the orthogonally equivalent matrix QAQ', where Q is the matrix of the eye rotation (Birkhoff & MacLane, 1965, p. 245).

Pursuit eye movements. Further simplifications of the parameters of image transformation may be produced if the eye tracks the movements of a certain image point, thus stabilizing it on the retina. Algebraically, pursuit movements change the transformation matrix A into the row-equivalent matrix QA, where Q is the matrix of the eye rotation (Birkhoff & MacLane, 1965, p. 245).

These observations on the effects of eye movement suggest that exploring the optical array with the eyes, and determining when the foveal transformation achieves a certain simplification, may serve as a visual-motor means for identifying the important parameters of the object displacement (cf. Gibson, 1959).

GENERAL DISCUSSION

This study has followed the program suggested by Gibson (1959) of looking for high-order variables of optical stimulation which are in close correspondence to the environment, and may therefore serve as sources of information to the visual system. It was shown that, as Gibson (1957) hypothesized, optical transformations are in close correspondence to object displacements, and also—as Gibson did not make clear—to their starting positions.

Two limitations of the present study should be noted: First, only object displacements, not changes of size or shape, were considered. If object deformations were also to be considered, the range of optical equivalents to an optical transformation would be much enlarged; the shadow deformation is itself one such equivalent (Figure 1A). In short, optical transformations contain information relative to a specific range of possibilities, the range of object displacements. It is reasonable to suppose that this range is of special significance. For one thing, object displacements are the class of object changes which can be matched, or reversed, by appropriate movements of the subject's head (cf. Fieandt & Gibson, 1959; Poincare, 1921). A second limitation on the present study was that neither the trajectory nor the speed of object displacement was considered; for purposes of the transformation analysis, these were unrestricted. But such displacement properties may well contain additional information, especially insofar as inertia prevents rapid changes in the axes of a displacement.

The present findings may be compared with those of other studies of the moving shadow illusion (Braunstein, 1962). Mach (1914) suggested that when an image is elongated or shortened along one dimension it induces the impression of a rotation in depth. Thus Mach identified one parameter of projective transformation, *stretching*, as critical. Wallach (1953) presented evidence which is in conflict with Mach's theory, and found that the critical stimulus for the impression of rotation in depth is the simultaneous change of length and angle in lines of the image; that is, shearing or foreshortening. Johansson (1964) found evidence that simultaneous change in both lateral and vertical image dimensions was the effective stimulus for motion in depth. The present analysis shows that any one of these hypothetical mechanisms would fail to accurately identify object characteristics except in very special conditions (p. 563). If the visual response is to be specific to the object characteristics, it must be governed jointly by all the parameters of projective image transformation: translation, stretching, shearing, and foreshortening.

REFERENCES

BIRKHOFF, G., & MACLANE, S. *A survey of modern algebra, 3rd edition.* New York: Macmillan, 1965.

BRAUNSTEIN, M. L. The perception of depth through motion. *Psychological Bulletin,* 1962, **59**, 422–433.

FIEANDT, K. VON, & GIBSON, J. J. The sensitivity of the eye to two kinds of continuous transformation of a shadow-pattern. *Journal of Experimental Psychology,* 1959, **57**, 344–347.

GIBSON, E., GIBSON, J. J., SMITH, O. W., & FLOCK, H. Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology,* 1959, **58**, 40–51.

GIBSON, J. J. Optical motions and transformations as stimuli for visual perception. *Psychological Review,* 1957, 64, 288–295.

GIBSON, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 456–501.

GIBSON, J. J., & GIBSON, E. Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology*, 1957, **54**, 129–138.

GIBSON, J. J., OLUM, P., & ROSENBLATT, F. Parallax and perspective during aircraft landings. *American Journal of Psychology*, 1955, 68, 372–375.

ITTLESON, W. H. *Visual space perception.* New York: Springer, 1960.

JOHANSSON, G. Perception of motion and changing form. *Scandinavian Journal of Psychology*, 1964, **5**, 181–208.

KLEIN, F. *Elementary mathematics from an advanced standpoint: Geometry.* New York: Dover, 1908.

MACH, E. *The analysis of sensations.* Chicago: Open Court, 1914.

METZGER, W. Tiefenerscheinungen in optischen Bewegungsfeldern. *Psychologische Forschung*, 1934, **20**, 195–260.

POINCARE, H. *The value of science.* New York: Dover, 1958.

SCHIFF, W., CAVINESS, J., & GIBSON, J. J. Persistent fear responses in rhesus monkeys to the optical stimulus of "looming." *Science*, 1962, **136**, 982–983.

WALLACH, H., & O'CONNELL, D. N. The kinetic depth effect. *Journal of Experimental Psychology*, 1953, **45**, 205–217.

(Received September 9, 1965)

Psychological Review
1966, Vol. 73, No. 6, 566–578

A FREQUENCY THEORY OF VERBAL-DISCRIMINATION LEARNING[1]

BRUCE R. EKSTRAND, WILLIAM P. WALLACE, AND BENTON J. UNDERWOOD

Northwestern University

The frequency theory of verbal-discrimination (VD) learning postulates that the cue for discrimination is the difference in frequency of occurrence between the correct (C) and incorrect (I) alternatives of a VD pair. It is postulated that, as VD learning proceeds, at least a 2:1 frequency difference in favor of the C item is built up, and Ss use this difference as their cue. By always selecting the item with the greater perceived frequency of occurrence, Ss can perform efficiently. An experiment was presented which produced results in support of the frequency theory. The discussion centered on additional predictions of the theory and the evidence available in support of these predictions.

The present paper deals with the processes involved in verbal-discrimination (VD) learning. In VD learning, pairs of verbal units are presented to the subject (*S*), and his task is to discover which unit in each pair has been arbitrarily designated by the experimenter (*E*) as being the correct unit. The correct alternative shall be referred to as the C unit and the incorrect alternative as the I unit. Normally a list consisting of several VD pairs is presented on a memory drum, and *S* must pronounce the correct response within a prescribed time limit. In a typical VD list, the C and I items are presented to *S* one above the other. (Half the time the C unit is presented above the I unit, and half the time the two are reversed.) The *S* is requested to choose one of the two units and pronounce it, after which *E* informs *S* as to the correctness of his choice. This information has been given verbally—*E* says "right" or "wrong"—or visually—the memory drum is programmed to indicate the C item after *S*'s choice is made. It has also been typical of VD procedures to omit a study trial at the beginning of learning; since there are only two alternatives (although more alternatives could be employed), *S* has a 50:50 chance of being correct without any learning taking place.

The theory to be presented here was first developed to explain the results of an experiment on transfer from one VD list to another VD list, but it became apparent that it could be used to interpret the manner in which a single VD list is learned. It will be useful to review first the results of that transfer experiment (Underwood, Jesse, & Ekstrand, 1964) and the manner in which the results were interpreted. The extension of the interpretation to single-list VD learning will then be considered.

TRANSFER EXPERIMENT

The Underwood et al. (1964) experiment employed three conditions. The control condition involved the learning of two unrelated VD lists. In the wrong condition (W), *Ss* learned one VD list to criterion and then transferred to a second VD list

---

[1] This work was supported in part by Contract Nonr-1228 (15), Project NR 154-057, between Northwestern University and the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose by the United States Government.

in which the I item in each pair in List 1 remained the I item in List 2, but the C items of List 1 were replaced with new C items in List 2. Condition R (right) was the reverse of Condition W; the C items of List 2 were the same C items that were in List 1, but all of the I items were changed from List 1 to List 2.

The results indicated that Condition R showed essentially 100% transfer. Transfer was initially almost as high in Condition W, but these *S*s showed a strange inability to improve their performance across trials. Indeed, the control group managed to reach criterion on List 2 before the W group.

Underwood et al. offered the following interpretation for their results. The basis for the discrimination in List 2 was the subjective difference in frequency of occurrence between the C and I items within a pair. In Condition R all of the I items were "new" items in List 2, and all of the C items were "old"—they had been in List 1. By always selecting the old items *S*s could perform perfectly, which they essentially did. Always selecting the more frequent of the two units in a pair is a "rule" for VD learning which shall be referred to as Rule 1. The term rule shall be used for convenience of discussion; this usage should not be construed as necessarily meaning that *S* consciously applies a rule in the literal sense of the word. In the W condition, all of the I items were old, and all of the C items were new. Therefore, the appropriate rule would be always to guess the less frequent of the two units in a pair (Rule 2). Since *S*s were fully informed about the relation between List 1 and List 2, and since at the start of learning of List 2 it should be very easy to tell new items from old, initial transfer would be high in both groups, and indeed it was.

It can be seen, however, that as trials continued on List 2 the new items would start to get old, and the old items would get older. Underwood et al. reasoned that C items get older faster than I items because the C items are pronounced once in the act of anticipation as well as seen once on the memory drum. The I items, on the other hand, are only seen on the drum and are not pronounced. Thus, each trial adds two frequency "units" to the C items and one unit to the I items. This is the basic postulate of the present theory.

Assuming this 2:1 frequency difference in favor of the C items, the failure of the W groups to improve with practice is understandable. The new C items get two units on each trial whereas the old I items are getting older, but only at the rate of one unit on each trial. At some later point in learning, the new C items will be as old as the I items, and beyond that point they will be older than the I items. Application of Rule 2 from this point on will result in an incorrect response because Rule 1 is now appropriate. For any given VD pair, Rule 1 or Rule 2 may be the appropriate one, depending upon the age of the C and I items. Neither rule could be applied with success until all of the C items had become older than their corresponding I items; therefore, the performance should suffer.

SINGLE-LIST VD SITUATION

The purpose of the present paper is to extend the basic hypothesis of Underwood et al., designed to explain transfer results, to the single-list VD situation. It is assumed that in learning a single VD list, the cue for discrimination is the subjective difference in frequency of occurrence between the C and I item in each VD pair. More particularly, it is asserted that as trials proceed there is built up at least a 2:1

frequency difference in favor of the C item, and that this difference can serve as the discriminative cue so that application of Rule 1 will result in the correct response.

Let us trace the sequence of events in the learning of a VD list and indicate the critical events assumed by the theory. On the guessing trial, a pair of items is presented, and *S* reads each one. This act of perception of each alternative has been referred to as a representational response (RR) by Bousfield, Whitmarsh, and Danick (1958). The RR to each item, according to the present theory, adds one frequency unit to the C item and one to the I item. After the two RRs have been made, *S* chooses one of the two items and pronounces it; this is his guess as to the correct one. This pronunciation response shall be referred to as the PR. It is assumed that the PR adds another frequency unit to whichever item is pronounced. If *S* made his PR to the C item, there would be a 2:1 frequency difference in favor of the C item. If *S* chose the I item, there would be a 2:1 difference in favor of the I item.

Now, by chance, half the PRs would be I items and half would be C items, so that if no frequency units other than the RR and PR were involved, the frequency-difference cue would be useless. The more frequent item in each pair would be correct in half the pairs (those pairs for which *S* guessed the C item) and incorrect in half the pairs (those for which *S* guessed the I item). Rule 1 would be appropriate for the former type of pair, and Rule 2 would be appropriate for the latter type of pair. In order to learn on this basis, *S* would have to learn which pairs go with Rule 1 and which go with Rule 2. For this reason, it is believed that a third type of response is occurring, namely, a rehearsal-of-the-correct-alternative response (RCR). This RCR may be thought of as a pronouncing of the C item implicitly or explicitly, and it may occur more than once immediately after *E* supplies the information about the correctness of *S*'s response.

The nature of the hypothesized processes in VD learning can now be more explicitly stated. In general, it is asserted that the single VD list of unrelated items is learned by application of Rule 1; namely, select the more frequent alternative. In order for this to be adaptive, Rule 1 must be appropriate for all the pairs; that is, the unit with greater frequency must be the correct one. Pairs for which *S* guesses the C item on the first trial will have a 2:1 difference in favor of the C item, so Rule 1 will be appropriate. Pairs for which *S* guesses the I item, as pointed out, would have a 2:1 difference in favor of the I item, so Rule 1 would be inappropriate. However, if *S* makes one RCR after having been informed that his guess was incorrect, the difference in frequency would be eliminated. Such pairs would now have 2:2 ratios, and on the next trial if *S* guessed and this time chose the C item, there would then be a 4:3 frequency ratio in favor of the C item. Now Rule 1 would be appropriate for this item. If *S* makes more than one RCR after choosing the I item, the pair would immediately become appropriate for Rule 1. Implicit in this discussion is the prediction that if the guessing trial were eliminated (thus eliminating some of the PRs to the I items), performance would be enhanced.

The extent to which these RCRs are made by *S*s is not known, but for frequency to be the appropriate cue and for Rule 1 to be the appropriate rule for every pair, it is only necessary to postulate that one RCR is made, and that it is made only when *S* guesses the

I item. In other words, the frequency unit that is added by the RCR just offsets the frequency unit that the I item received as a result of the PR. This returns the pair to a "neutral" state (2:2) so that there is a 50:50 chance that *S* will make his PR to the C item on the next trial. If *S*s make RCRs more than once for any given pair, and if they make RCRs when they have chosen the C item, the frequency cue would only be further enhanced. For example, if *S* makes a correct choice on the guessing trial, it has been postulated that there is a 2:1 difference in favor of the C item. But if, in addition, *S* makes one RCR, there would be a 3:1 difference.

As trials continue, there will be a buildup in frequency units for the C and I items, but this buildup will always tend to favor the C item. As a result, the difference in frequency units between the C and I items will increase, making the discrimination between the more and less frequent items easier. Note that the interpretation states that the number of frequency units for each item can be determined by counting the RRs, PRs, and RCRs that have been made to that item. It also is apparent from the above discussion that it has been assumed that one RR adds as much to the subjective frequency of an item as does one PR, or RCR; that is, the equality of RRs, PRs, and RCRs has been assumed. Both of these assumptions are merely working hypotheses. It is quite possible that the three postulated responses are unequal, that an RCR is "worth more" than an RR, or that a PR of the C item (a correct response) is somehow qualitatively different than a PR of the I item (an error) in terms of the consequences for frequency units. For simplicity's sake it has been assumed that these responses are equal in terms of frequency units added.

A fourth way in which frequency units may be added to VD items is through an implicit associative response (IAR). Bousfield et al. (1958) hypothesized that the presentation of a verbal unit elicits IARs related to the given verbal unit. Thus, if the word DOG is presented, an IAR of *cat* (or some other related item) is presumed to occur. In a VD list the frequency of a given unit may be increased if that unit has occurred elsewhere in the list as an IAR. For example, if the word *cat* appears in the list, it will attain a certain frequency value after one presentation based on RRs, PRs, and RCRs. The appearance of the word DOG in the same list may increase the frequency value of the word *cat* to the degree that DOG elicits *cat* as an IAR. To the extent that the associations are bidirectional, the appearance of CAT may add frequency units to *dog*. In other words, *S* may "confuse" an IAR with an RR, so that these IARs may add frequency units in the same sense that an RR does. This frequency alteration by an IAR mechanism is assumed to occur only in specific instances (e.g., VD lists containing associatively related words).

Another major process postulated here is that *S*s can and do discriminate items on the basis of their frequency of occurrence. That they can make accurate judgments of the frequency of an item in the language has been shown by Underwood (1966). The Underwood et al. (1964) transfer study shows that when frequency differences are extremely large, performance is essentially perfect; a fact which strongly suggests that the cue for discrimination in that study was the frequency difference. It has also been stated that *S*s learn single VD lists by application of Rule 1; however, there is no reason to believe that when the frequency difference is reversed (i.e., the less frequent member of each pair is correct), *S*s cannot operate using Rule 2. The W condition of the Underwood et al. (1964) study shows that they can. Thus, either Rule 1 or 2 can be applied depending on the appropriateness of the situation. In the single-list VD situation where *S* is required to pronounce the C item, there is a built-in difference in frequency, due to this PR, which makes Rule 1 more appropriate.

THE EXPERIMENT

The frequency theory, first developed to explain the transfer results in the Underwood et al. (1964) experiment, was admittedly postulated after the data were collected and analyzed. As such, the theory requires verification under other situations, and this was the purpose of the present experiment. More particularly, it was desired that the theory as applied to the single-list VD situation be tested.

It was reasoned that if the discriminative cue is the relative frequency of the two units in a pair, increasing the difference in frequency should facilitate performance (a difference of 4:1 should be easier than 2:1), and decreasing the difference should inhibit performance (a "difference" of 2:2 should be more difficult than 2:1). The method decided upon to produce these increases or decreases was to have an item occur in two pairs; therefore, the repeated item would be doubled in frequency, receiving two RRs per trial instead of one. For the moment, let us consider only the frequency units added by RRs and PRs, and let us assume that the frequency ratio is normally 2:1 when there are no repeated items in the list.

In the first condition, the same item occurs in two pairs, and it is correct in both pairs; this will be referred to as the same-right condition (SR). Since the repeated item is the C item in the two pairs in which it occurs, it will have four frequency units added to it on each trial (two RRs and two PRs, assuming *S* guesses the C item). The I item in each of those pairs will have only one unit added on each trial (an RR). Thus, for each pair there will be a 4:1 ratio in favor of the more frequent item, and application of Rule 1 will result in the correct response. If a 4:1 ratio makes for easier discrimination than a 2:1 ratio, Condition SR should surpass a control condition in which there are no repeated items in the list.

In the second condition, the same item occurs in two pairs, but it is incorrect in both pairs; this was called the same-wrong condition (SW). Since the I item gets only one unit per trial (an RR), the repeated item in this condition will now have only two units added to it on each trial (one RR for each pair in which it occurs). The C item in each of these pairs will also get two units added per trial (one RR and one PR). Thus, the ratio is 2:2 when only RRs and PRs are considered. This should be considerably more difficult than the SR condition where the ratio is 4:1 and also more difficult than the control condition (2:1).

The third condition investigated was designed to produce maximum interference by making the frequency cue very difficult to use. In this condition the repeated item is the C item in one of the two pairs in which it occurs and the I item in the other pair; this shall be referred to as the same-both condition (SB). Consider the pair in which the repeated item is correct; it receives two units for being correct (one RR and one PR) and one unit from the fact that it is repeated in another pair (an RR). The I item in this pair receives only one unit, an RR. The resultant ratio is 3:1 with the more frequent one being correct. However, in the pair in which the repeated item is incorrect, the situation is reversed. The repeated item has a frequency of three (an RR and a PR from the other pair, and an RR from this pair), and the C item in this pair has a frequency of two (one RR and one PR). The resultant ratio is 2:3 with the *less* frequent item being correct. As a result, Rule 1 would be appropriate for the pair in which the repeated item is correct, but Rule 2 would apply to the pair in which the repeated item was incorrect. Such an arrangement should produce considerable interference, so the SB group should be inferior to the other three conditions, SW, SR, and the control. In summary, the frequency theory predicts that the SR group should surpass the control condition which should surpass the SW condition. Finally, the SB group should do poorest.

As a second method of increasing or decreasing the frequency ratios within a VD pair, strong associates of some of the items in other pairs were inserted. Thus, one pair might be *Queen-Fast* and another pair, *King-Pepper*. Instead of repeating *Queen* to add to its frequency, we reasoned that presentation of *King* would lead to an IAR which would most likely be the word *Queen*. Of course, to the extent that the associations are bidirectional, presentation of *Queen* would also lead to an IAR of *King*. Underwood (1965) has independent evidence that *S*s may confuse the RR to an item with the same response given as an IAR to a strong associate of that item. For example, if *King* were actually presented, but *Queen* was never presented, *S*s would indicate, nevertheless, that *Queen* (the IAR to *King*) had actually been presented. The present manipulations were designed to explore the possibility, mentioned earlier, that IARs may contribute frequency units in the same sense that RRs, PRs, and RCRs are hypothesized to increase frequency.

TABLE 1

SAMPLE PAIRS

Control	Same Right (SR)	Same Wrong (SW)
Queen[a]Fast Door[a]-Pepper	Queen[a]-Fast Queen[a]-Pepper	Queen-Fast[a] Queen-Pepper[a]
Same Both (SB)	Associate Right (AR)	Associate Wrong (AW)
Queen[a]-Fast Queen-Pepper[a]	Queen[a]-Fast King[a]-Pepper	Queen-Fast[a] King-Pepper[a]
	Associate Both (AB)	
	Queen[a]-Fast King-Pepper[a]	

[a] Denotes the correct item in each pair.

Using strong associates, the three experimental conditions described above were repeated. In Condition associate right (AR), the strong associates were the C items in their respective pairs; in Condition associate wrong (AW), the associated items were the I items in their respective pairs; in Condition associate both (AB), one item would be the C item in its pair, and the associate of that item would be the I item in its pair. To the extent that *S*s make IARs that are words in other pairs, the predictions for these associated (A) conditions are the same as those for the conditions in which an item is actually repeated (S conditions). The AR condition should produce facilitation, the AB condition should produce interference, and the AW condition should fall in between AR and AB, and below the control condition. Examples of the seven types of pairs are presented in Table 1.

Design

The design was a 3×2 factorial, with 2 types of interpair relationships (S or A) and 3 types of lists (R, W, or B). This

TABLE 2

LISTS USED

Control list	S list	A list
Cold-Sickness	Cold-Sickness[a]	Cold-Sickness[a]
Boy-Woman	Cold-Woman	Hot-Woman
Chair-Head	Chair-Head	Chair-Head
Carpet-Foot	Chair-Foot[a]	Table-Foot[a]
House-Rough	House-Rough[a]	House-Rough[a]
Dark-Bread	House-Bread	Home-Bread
Tiger-Wool	Tiger-Wool	Tiger-Wool
Nail-Short	Tiger-Short[a]	Lion-Short[a]
Queen-Fast	Queen-Fast	Queen-Fast
Door-Pepper	Queen-Pepper[a]	King-Pepper[a]
Needle-Sweet	Needle-Sweet[a]	Needle-Sweet[a]
Ugly-Shallow	Needle-Shallow	Thread-Shallow
Road-Web	Road-Web	Road-Web
Round-Doctor	Road-Doctor[a]	Street-Doctor[a]
Hill-Smoke	Hill-Smoke	Hill-Smoke
White-Baby	Hill-Baby[a]	Mountain-Baby[a]
Steal-Soft	Steal-Soft[a]	Steal-Soft[a]
Loud-Apple	Steal-Apple	Thief-Apple
Town-Low	Town-Low[a]	Town-Low[a]
Bed-Cat	Town-Cat	City-Cat

Note.—In the Control list, the pairs are given in the order: C item—I item; Cold was correct, Sickness was incorrect. In the A and S lists, both of the associated or repeated items were correct in conditions AR and SR, incorrect in AW and SW, and one was correct, one incorrect in AB and SB.

[a] Indicates those pairs in which the associated or repeated item was incorrect in the AB or SB conditions, respectively.

produces the 6 experimental groups described above: SR, SW, SB, AR, AW, and AB. In addition, a control group was run. The *S*s in this group learned a VD list in which there were no formal interpair relationships.

The frequency theory predicts a main effect of type of list—the R conditions should surpass the W conditions, both of which should surpass the B conditions—and an interaction of type of List × type of Interpair Relationship. This interaction is predicted because the effect of type of list (R, W, or B) should be greater when the same item occurs in two pairs (S conditions) than when strong associates occur in different pairs (A conditions). Repeating the same item will always add one frequency unit, but presentation of a strong associate will add a frequency unit only if *S* makes the appropriate IAR.

Lists

Each of the seven lists contained 20 VD pairs. The items were selected from the Russell and Jenkins (1954) and the Palermo and Jenkins (1964) word-association norms. From these norms 40 pairs of associated items were selected; it was attempted in the process to keep interpair associations at a minimum. By selecting at random one word from each pair, the 40 presumably unrelated words for the control list were obtained. These 40 words were paired randomly to make the 20 VD pairs, and within each pair one word was randomly designated as the C item for that pair; the remaining item, of course, became the I item.

In order to construct the experimental lists 10 pairs were chosen at random from the 20 control pairs. In addition, the 10 I items from the remaining 10 control pairs were used. This provided a pool of 10 pairs and 10 single items. From each of the 10 pairs chosen from the control list, two pairs for use in the experimental lists were constructed. For example, one control pair was *Queen-Fast;* in the S list this became two pairs, *Queen-Fast* and *Queen-Pepper, Pepper* having been an I item in one of the 10 control pairs not used. In the A list, the pair *Queen-Fast* also was used to make up two pairs, *Queen-Fast* and *King-Pepper.* By using the 10 control pairs and the 10 I items selected from the control list, 20 experimental pairs were constructed with maximum overlap between S, A, and control lists. The pairings for the S, A, and control lists are presented in Table 2.

Once the pairs were determined, the right and wrong members were designated as follows: In the AR list, the two associated items (e.g., *Queen* and *King*) were both correct in their respective pairs; in the AW list, the pairs were simply "turned over" so that the two associated items were wrong in their respective pairs (i.e., *Fast* and *Pepper* were correct). In the AB list, one of the two related pairs (chosen at random) was turned over so that the I item in one pair was a strong associate of a C item in another pair (e.g., *Queen-Fast, Queen* is correct; *King-Pepper, Pepper* is correct). For the S lists the procedure was identical; in SR, the repeated item was correct in both pairs in which it occurred; in SW, the list was turned over so that the repeated item was incorrect in both pairs; in SB, the repeated item was correct in one pair but incorrect in the other pair in which it appeared.

Procedure

The procedure employed group learning sessions with the materials being presented via a tape recorder. The learning method was alternate study and test trials with a modified anticipation procedure on the test trials. Four study and four test trials were administered to each of the seven groups. No guessing trial was used. The presentation rate was 3 seconds for both study and test trials.

On study trials, the two items in each pair were read, and then the voice on the tape repeated the correct one; for example, "*Queen-Fast, Queen* is correct." Both the reading of the pair and the saying of the C item took place within the 3-second interval. This procedure insured that a frequency difference in favor of the correct alternative existed, because the C item was said twice on the study trials. On test trials, the pairs were again read at a 3-second rate, but the voice did not indicate correctness. The *S*s recorded their responses on IBM answer sheets numbered from 1 to 80, with two alternatives for each of the 80 numbers. As each pair was read, *S*s indicated in the corresponding place on their answer sheets which member of the pair (first or second) was correct. Since the nature of this procedure does not allow *S* to make a PR, the frequency unit for the PR is eliminated. Instead, the voice on the tape spoke the C item twice on study trials to insure the 2:1 ratio postulated by the theory. Across the four study and four test trials, the C item of each pair was read first four times and second four times. Within a trial, the ordering of the C and I items was random. Eight random orders of the 20 pairs were used for the eight presentations.

Subjects and Instructions

The *S*s were college students from introductory psychology courses at Northwestern University; there were 20 in each of the seven groups. The seven conditions were each run twice, so a total of 14 group sessions was administered. These 14 sessions were run at either 1:00 P.M. or 1:30 P.M. on 7 consecutive school days. The *S*s made an appointment for one of the 14 sessions at their convenience. The seven conditions were assigned at random to the 14 different sessions, subject to the restriction that each condition occur once before any condition received its second session. The result was that there were 20, 21, or 22 *S*s in each of the seven groups. IBM sheets were randomly discarded until each group consisted of 20 *S*s.

The *S*s were read ordinary instructions for VD learning with the necessary modifications for the group procedure. No mention was made of the possibility of repeated or associated items.

RESULTS

The results (mean number of total errors) are shown in Figure 1—acquisition curves revealed no interactions with trials. The base line in Figure 1 represents an assumed continuum of similarity of implicit responses ranging from unrelated items (control) to associated items (A) to identical or same items (S). The plotting parameter is the type of list, either R, W, or B. In addition, for convenience of comparison, a dotted line representing the control mean is included.

An initial analysis of variance was performed on all seven groups. This produced an F (6, 133) of 25.68 which is highly significant. This indicates that the different treatments did produce differences in learning. Next a 3 × 2 analysis of variance was performed on the six experimental groups. The results indicated that both main effects and the interaction were significant at better than the .01 level. The main effect of A versus S interpair relation-

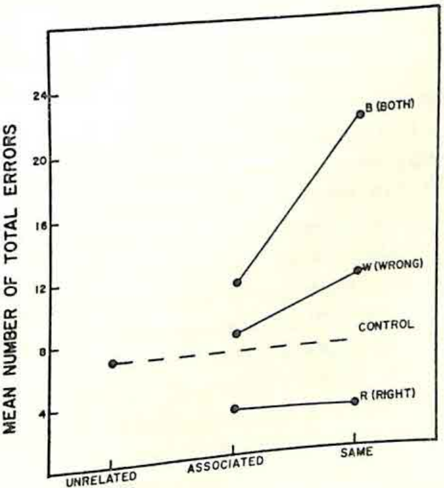

FIG. 1. Mean total errors in four trials of VD learning for the six experimental groups and the control group. (The dotted line represents the control mean and is presented for the convenience of comparison.)

TABLE 3

Mean Total Errors for the Seven Groups and Results of the Duncan Multiple-Range Test

SB	SW	AB	AW	Control	AR	SR
21.50	11.35	11.30	8.00	6.90	3.05	2.80
					———	———
			———	———		
	———	———	———			

Note.—Any two means underlined by the same line are not significantly different (Edwards, 1960).

ship produced an F (1, 133) of 18.48, that of type of list (R, W, or B) an F (2, 133) of 52.78. The interaction gave an F (2, 133) of 8.83. It can be seen from the *df* that these Fs are based on the within-groups term for all seven groups combined.

The results indicate that the A lists were easier than the S lists. Second, the type of list (R, W, or B) is important: The R conditions are clearly superior to the W conditions which, in turn, are superior to the B conditions. The interaction indicates that, as predicted, these differences between R, W, and B lists are more pronounced with the S interpair relationship than with the A relationship.

The control-group mean falls exactly as predicted, between the R and W conditions. Thus, numerically, the W and B conditions show interference, and the R conditions show facilitation. To evaluate the differences among the seven groups, a Duncan multiple-range test was performed using the .05 level of significance (Edwards, 1960). The results are shown in Table 3 where the seven means are given.

Within the S conditions, SR is significantly superior to the control condition whereas SW and SB are significantly inferior to the control. These four groups are thus ordered and differ significantly as predicted by the theory. Furthermore, since the SW group differs from the control group, interference in this group has apparently been produced by repeating an I item in another pair.

Within the A conditions, the results are not as conclusive as for the S conditions. The four groups (including the control) are ordered in the same way as in the S conditions, but, as expected, the differences among them are considerably smaller. The AR group is significantly superior to the control group, and AB is significantly inferior to the control, indicating facilitation and interference, respectively. The AW group, however, is only slightly and insignificantly inferior to the control group, which may indicate that there is no interference from a strong associate being in the list when both associated items are I items.

Discussion

The results of this experiment provide substantial support for the frequency theory. Both main effects and the interaction in the 3 × 2 design were significant at the .01 level in the predicted direction. Within the A conditions and within the S conditions, the groups were ordered exactly as predicted. By increasing the difference in frequency between the C and I items (R conditions), performance was improved, and by decreasing this difference (W conditions), interference, although small in amount, was apparent. Much more striking inhibition was apparent in the B conditions which were designed to produce maximum difficulty for frequency discrimination.

The results would also tend to support the idea that these frequencies can be manipulated by presenting associates of some items in other pairs. However, due to the small effects of this operation, a strong conclusion is not indicated. Nevertheless, it appears that under optimal conditions, the perceived frequency of the word *Queen* can be increased by presenting the word *King*. This is taken as further support for the operation of an implicit associative-response process in *S*s as they respond to verbal stimuli.

GENERAL DISCUSSION

The above experiment has put the frequency theory to a rather severe test, and it has survived. In view of this success, other predictions of the theory and related evidence in the literature shall be examined.

Similarity

Increasing intralist similarity should interfere with VD learning (the similarity existing between correct and incorrect alternatives). The more similar two items are, the more likely an RR, PR, or RCR made to one item will be confused with an RR, PR, or RCR made to another item. The SB condition from the experiment reported here is just an extreme case of this similarity (identity), and SB was shown to be an interference condition. Both Underwood and Archer (1955) and Battig and Brackett (1963) have shown that increasing formal intralist similarity in a VD list can produce interference.

Underwood and Viterna (1951) compared VD lists of high and medium meaningful similarity manipulated within a pair, keeping interpair similarity low. Thus, the two adjectives within a pair were either of high or medium synonymity. Their results indicated a small but significant difference in favor of the medium-similarity group. Although the interference in the present experiment was entirely interpair interference, with intrapair similarity at a minimum, the frequency theory would predict interference in either case. If *King* and *Queen* are in the same pair, the IAR to *King* would tend to increase the frequency of *Queen*, and vice versa, just as if *King* and *Queen* were in two different pairs. Thus, the Underwood-Viterna conditions were quite similar to the AB condition which showed evidence of interference.

Meaningfulness

The theory would seem to require the prediction that, with integrated units, meaningfulness of the units (in terms of frequency) should have no effect. If the two units in a VD pair are of equal frequency, whether that frequency is high or low on an absolute scale, adding an additional frequency unit to either item should make the pair more discriminable. The theory might even predict (according to Weber's law) that adding one unit to an item already high in frequency should be less noticeable than adding one frequency unit to an item low in frequency. This would mean that VD pairs of very-high-frequency units would be more difficult than pairs of low-frequency units. Assuming that frequency units established outside the laboratory are comparable to frequency units established by RRs, PRs, and RCRs in the laboratory, one might then predict that VD pairs of high-frequency words would be more difficult than pairs of low-frequency words. However, the applicability of Weber's law in this situation has not been demonstrated, and neither has the equality of experimental and extraexperimental frequency units.

There is some evidence that VD learning is not strongly affected by the meaningfulness of the items. Keppel (1966) found no difference in trials to criterion between a VD list of nonsense syllables and a list of words. Postman (1962) found that VD pairs of high-frequency words were more difficult than pairs composed of either low- or medium-frequency words. It should be pointed out, however, that the difficulty of the high-frequency list may be a result of a greater number of interitem associations in that list (Deese, 1960). To the extent that there are interitem associations in Postman's high-frequency list, it is similar to the AB list of the present experiment (an interference condition) which may explain why it was more difficult than the medium- and low-frequency lists.

If a detrimental effect of low meaningfulness is found for VD learning, the theory would predict that this would be with nonintegrated units in which the RRs, PRs, and RCRs are letter responses, instead of responses to the entire unit. To the extent that there are repeated letters in the list, the frequency theory would predict interference, since the RR to a nonintegrated unit could easily be confused with the RR to another unit with similar letters (e.g., Underwood & Archer, 1955). Moreover, the lower the integration, the less reliable the RR is likely to be; that is, the less similar two RRs to the *same* unit will be. Until the unit is integrated sufficiently for RRs to be reliable, we cannot expect a buildup of a particular PR that can be used as a frequency cue.

The theory need not necessarily predict a difference in VD learning as a function of meaningfulness or frequency of occurrence of words in the language since extraexperimental frequency may not be comparable to experimentally induced frequency. However, if frequency differences are induced entirely within the experiment (possibly by familiarization training), it is possible that a VD pair of two high-frequency units will be more difficult than a VD pair of two low-frequency units. Other things being equal (interitem associations in particular), an extra frequency unit added to a member of a high-frequency pair should not be "worth as much" in terms of cue value as a frequency unit added to a member of a low-frequency pair.

Presentation Rate

The theory makes a rather unusual prediction with regard to the effects of anticipation time. A large increase in anticipation time, with study time held constant, should have no effect on VD performance, or perhaps even an inhibitory effect. If *S* is forced to spend extra time studying the two units in a VD pair before he is allowed to make his choice, performance could suffer. This additional time should result in additional RRs being made to the C and I items. If the same number of RRs is added to the C and I items by this extra time, the relative frequency difference can only decrease, making the discrimination more difficult (again assuming the applicability of Weber's law). An analogy can be made with the strategy often suggested for taking multiple-choice examinations; namely, "stick with your first choice." Continued study of the alternative answers should tend to reduce their discriminability. On the other hand, if the study time after reinforcement is increased, the theory predicts better performance, on the assumption that this time will be spent mainly in the act of making RCRs. Increasing the study time after reinforcement should increase the number of RCRs relative to the number of RRs to the I item; any such increase should result in a more discriminable frequency difference.

Forgetting

If one subscribes to a strict interference theory of forgetting and applies it to the frequency theory, one would have to predict no forgetting of a VD pair unless there were "occurrences" of the C and I items during the retention interval. Such occurrences would add frequency units to both items and reduce the discriminability of the pair. By appropriate manipulations in a second VD list, it should be possible to produce a variety of positive or negative transfer effects (e.g., Underwood et al., 1964) and either retroactive inhibition or facilitation when recall of the first list is measured. Of course, in addition to the interference effects, there may be some loss of the frequency cue due to a decay process.

From the standpoint of a strict interference interpretation, the theory must predict that increasing the number of pairs in a VD list will have no effect on the number of trials required to learn that list. This prediction, however, will hold only if it is possible to increase list length without also increasing the number of interitem associations. Such an increase in interitem associations would tend to make the longer lists analogous to the AB list of the present experiment and would result in poorer performance.

Further Implications

Some predictions based on the frequency theory and some findings of other investigators which may be interpreted within the framework of the theory have been noted. The fact that other theoretical approaches may be developed which will give an equally good accounting of the data available is fully recognized. At the same time, however, it must be emphasized that there is a critical finding which must be handled by any alternative theoretical approach. This finding, predicted by the frequency theory, showed a large difference in learning between a list in which repeated items were correct and a list in which repeated items were incorrect. That learning of the latter list resulted in some interference (when compared with a control) only adds to the problem with which other theories must cope.

The assumption that differences in frequency of events provide a discriminative cue should have applicability to phenomena beyond those associated with VD learning. Some possibilities for such extensions of the theory may be mentioned briefly. It has often been noted that in tasks where verbal units must be recalled (paired associates, free learning) it is a rare occurrence for *S* to produce a unit that is not in the list. Even if the response units are 10 single letters drawn randomly from the 26 letters, *S* will rarely produce one of the 16 letters not in the list. It is as if some selector mechanism (Underwood & Schulz, 1960) quickly and effectively differentiates between the appropriate and inappropriate pool of responses. A frequency discrimination may be responsible for this selective capacity. The letters in the list have a situational frequency of at least one each, those not in the list have none, thereby providing the discriminative cue. Even if *S* cannot recall response terms which are in a list, he is able to reject response terms which are not in the list.

In learning the second list of certain transfer paradigms (e.g., A-B, A-B′), an appreciable number of intrusions from the first list may be observed in learning the second list (Postman & Stark, 1964). Furthermore, many of these intrusions appear after several trials on the second list (Schwenn & Underwood, 1965). This might be interpreted to mean that intrusions are most likely to occur when the associative strengths of two responses elicited by the same stimulus become equivalent. However, the frequency-discrimination theory would predict this phenomenon on the grounds that after the second-list responses have occurred several times the frequency discrimination between first- and second-list responses breaks down as the frequency of the second-list responses approximates that of the first.

Intrusions at recall in retroactive-inhibition studies decrease sharply in frequency as the degree of interpolated learning increases beyond the level of original learning (Melton & Irwin, 1940), although amount of retroactive inhibition does not. The reduction in intrusions would be expected from the frequency-discrimination theory since the higher degrees of interpolated learning make the frequency discrimination between first- and second-list responses easier.

REFERENCES

BATTIG, W. F., & BRACKETT, H. R. Transfer from verbal-discrimination to paired-associate learning: II. Effects of intralist similarity, method, and percentage occurrence of response members. *Journal of Experimental Psychology,* 1963, **65**, 507-514.

BOUSFIELD, W. A., WHITMARSH, G. A., & DANICK, J. J. Partial response identities in verbal generalization. *Psychological Reports,* 1958, **4**, 703-713.

DEESE, J. Frequency of usage and number of words in free recall: The role of association. *Psychological Reports,* 1960, **7**, 337-344.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart & Winston, 1960.

KEPPEL, G. Association by contiguity: Role of response availability. *Journal of Experimental Psychology,* 1966, **71**, 624-628.

MELTON, A. W., & IRWIN, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology,* 1940, **53**, 173-203.

PALERMO, D. S., & JENKINS, J. J. *Word association norms grade school through college.* Minneapolis: University of Minnesota Press, 1964.

POSTMAN, L. The effects of language habits on the acquisition and retention of verbal associations. *Journal of Experimental Psychology,* 1962, **64**, 7-19.

POSTMAN, L., & STARK, K. Proactive inhibition as a function of the conditions of transfer. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 249-259.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Technical Report No. 11, 1954, University of Minnesota, Contract N8 onr-66216, Office of Naval Research.

SCHWENN, E., & UNDERWOOD, B. J. Simulated similarity and mediation time in transfer. *Journal of Verbal Learning and Verbal Behavior,* 1965, **4**, 476-483.

UNDERWOOD, B. J. False recognition produced by implicit verbal responses. *Journal of Experimental Psychology,* 1965, **70**, 122-129.

UNDERWOOD, B. J. *Experimental psychology.* (2nd ed.) New York: Appleton-Century-Crofts, 1966.

UNDERWOOD, B. J., & ARCHER, E. J. Studies of distributed practice: XIV. Intralist similarity and presentation rate in verbal-discrimination learning of consonant syllables. *Journal of Experimental Psychology,* 1955, **50**, 120-124.

UNDERWOOD, B. J., JESSE, F., & EKSTRAND, B. R. Knowledge of rights and wrongs in verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior,* 1964, **3**, 183-186.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

UNDERWOOD, B. J., & VITERNA, R. O. Studies of distributed practice: IV. The effect of similarity and rate of presentation in verbal-discrimination learning. *Journal of Experimental Psychology,* 1951, **42**, 296-299.

(Received September 27, 1965)

Psychological Review
1966, Vol. 73, No. 6, 579–585

THEORETICAL NOTES

RESPONSE DISORGANIZATION AND BREADTH OF OBSERVATION IN SCHIZOPHRENIA

WILLIAM E. BROEN, JR.

University of California, Los Angeles[1]

2 variables: (a) degree of organization of competing responses; and (b) breadth of observation, are related to each other and used to discuss research tasks on which differences have been found between normals, acute schizophrenics, and chronic schizophrenics. Research results which have seemed contradictory can be clarified by the basic propositions that: (a) response disorganization characterizes both acute and chronic schizophrenics; and (b) chronics attempt to cope with response disorganization by observing fewer stimuli.

Consider these statements, each of which has good research support:

1*a*. Overinclusive concepts have been found in chronic schizophrenics (Chapman, 1961).

1*b*. Overinclusive thinking is found in many acute schizophrenics, but chronic schizophrenics are not abnormally overinclusive (Payne, 1962).

2*a*. Chronic schizophrenics attend to a narrower range of stimuli than normals (Venables, 1964).

2*b*. Chronic schizophrenics are more distractible than normals (Chapman, 1956).

When faced with such inconsistency we can either seek further data that will favor some conclusions and oppose others, or we can search for clarifying distinctions, and then use these distinctions to guide later research. The reliability of the data concerning the four statements listed above is such that it seems worthwhile to follow the latter course and think through the reasons for what may be only apparent inconsistencies. Discussion will be based on two processes: (*a*) response disorganization, and (*b*) narrowed observation.

RESPONSE DISORGANIZATION

Response disorganization refers to a lack of clear hierarchical ordering of competing response tendencies; that is, no single response tendency is highly dominant. The first principle to be used in the discussion is that many schizophrenics, both acutes and chronics, have abnormal response disorganization. More specifically, in a situation where a number of alternate response tendencies are evoked, in normal persons these response tendencies are hierarchically ordered with appropriate responses clearly dominant, but for disorganized schizophrenics the probabilities of the alternate responses are more nearly equal. Interference from competing responses is therefore increased. The increase in intrusions of responses which normally have low probability is seen as a major characteristic of schizophrenia. This principle, which is basically similar to Cameron's (1938) description of schizophrenic thinking, has been discussed in detail together with supporting evidence by Broen and Storms (1966).

It is quite likely that not all schizophrenics have abnormal response disorganization. One way of estimating the extent of this process comes from the suggestion by Broen and Storms (1966) and Maher (1966) that the widely noted tendency of many schizophrenics toward overinclusive thinking may be a consequence "of a more fundamental process by which all or many of the possible associations to a word are at equal strength [p. 413]." If this suggestion is accurate then the extent of overinclusive thinking may provide an estimate of the extent of the underlying process, which is here called response disorganization. As will be discussed later, some measures of overinclusive thinking may be influenced by more than one underlying process, especially in chronic schizophrenics. However Payne's (1966) statement that overinclusion ". . . is found in at most 50 per cent of a typical sample of acutely ill mental patients who are diagnosed as schizophrenic [p. 95]" may provide an estimate of the extent of response disorganization in acute schizophrenics. Although lack of complete uniformity must be kept in mind, research done on the heterogeneous groupings "acute schizophrenics" and "chronic schizophrenics" will be discussed as relevant material because the effects of the processes being discussed should make group averages differ in the expected directions.

---

[1] This paper was written while the author was a Visiting Lecturer at the University of Massachusetts, Amherst. Robert H. Harrison and Morton G. Harmatz gave helpful comments on earlier drafts.

To understand the effects of response disorganization, it is important to note that the different response tendencies, which many schizophrenics cannot organize adequately, may be evoked in two ways. First, a single stimulus may elicit several alternate responses. An example is a word in a word-association task which elicits several different associations. A good example of reduced hierarchical organization among alternate associates to a single stimulus was given by a schizophrenic:

> My trouble is that I've got too many thoughts. You might think about something, let's say that ashtray and just think, oh! yes, that's for putting my cigarette in, but I would think of it and then I would think of a dozen different things connected with it at the same time [McGhie & Chapman, 1961, p. 108].

Second, competing responses may be evoked by different stimuli. Consider what happens when a person enters a multiple stimulus situation. A number of different stimuli are observed, each presumably evoking a tendency to focus on that stimulus. Normally these competing focusing responses would be hierarchically ordered in accordance with learning and/or instructions; the strongest focusing tendencies evoked by the kinds of stimuli which are most important for appropriate performance. (Harlow, 1959, had discussed his research on learning set in similar terms.) Response disorganization would mean that these competing attention tendencies would be less clearly organized, that is, attention would be more scattered in schizophrenics than in normals. As stated by Shakow (1962):

> It is as if, in the scanning process which takes place before the response to a stimulus is made, the schizophrenic is unable to select out the material relevant for optimal response. He apparently cannot free himself from the irrelevant among the numerous possibilities available for choice [p. 9].

Attention to irrelevant stimuli increases the likelihood that additional competing response tendencies are evoked. So the problem is compounded for the disorganized schizophrenic. He not only has abnormal variation among competing response tendencies, but also, because of his scattered attention, he is likely to have a greater number of competing response tendencies. Exceptions to this statement will be discussed in the next section.

NARROWED OBSERVATION

The second major principle is that some schizophrenics, primarily nonparanoid chronics, learn to reduce interfering responses through reduced observation. We have discussed lack of organization among competing response tendencies evoked by different stimuli as one major source of interfering responses in schizophrenia. Reduced observation should tend to reduce this source of response confusion. Reduced observation is a process which would be learned over time and is therefore expected primarily in chronics. Reduced observation would also be learned primarily by those schizophrenics who most need reduced response interference, and for whom reduced vigilance is most comfortable. It is because there is some evidence suggesting that paranoids may be one group that is less disorganized than other schizophrenics (Lester, 1960),[2] and because paranoids would not be as comfortable with reduced vigilance, that reduced observation is expected primarily in nonparanoid chronics. New situations will be examined by these chronic schizophrenics. It is just that their breadth of observation is not as great. Reduced observation would, of course, decrease response complexity only when the alternate responses were stimulated by different stimuli, not when they are responses to a single stimulus which is observed.

Following Broadbent (1958), it is hypothesized that observation is not reduced randomly. Instead, observation is guided along the lines of discriminable stimulus classes (e.g., classes based on color, position, or stimuli from a single sensory modality), with some classes ignored and others observed normally. The probability that any particular stimulus class will be noticed is a function of such stimulus characteristics as: physical intensity (Broadbent, 1958); central position; and movement instead of being stationary, which will be called stimulus "pull." The relationship between stimulus pull and stimulus relevance is then quite important in determining the quality of chronic schizophrenics' performance in any particular task. In discussing past research, stimulus "pull" will be defined in terms of what would be commonly expected; for example, other things being equal, stimuli which are central in a stimulus display have stronger "pull" than stimuli which are peripheral. However, future research could use better definitions. One adequate definition would be obtained by showing the research situation to a reference group from the same population as an experimental group, having instructed the reference group to "tell me (or point to, etc.) what you notice first." The frequencies that different stimulus classes are indicated could define relative stimulus pull.

There is considerable research support for the idea that chronic nonparanoid schizophrenics use a subnormal range of stimuli (reviewed in Venables, 1964),[3] although whether the reduction is early in the attention process, as discussed here, or occurs later, is an open question. The question of causality is also more speculative than the principle that reduced observation does occur. The hypothesis that reduced observation is an attempt to cope with an increase in response interference by reducing the sources of stimulation is similar to ideas in discussions of schizophrenia by Freeman (1960), and Zahn, Rosenthal, and Shakow (1963). This notion that narrowed observation is a reaction to response interference is supported by Bursill's (1958) results indicating that, in a complex visual signal monitoring task, observation tends to be narrowed to more central stimuli when perceptual load is increased. (Bursill's research used normal subjects. Chronic schizophrenics are hypothesized to have learned more constant and extreme reductions than the temporary response to certain situations seen in normals.)[4]

---

[2] This does not mean that schizophrenics with paranoid delusions are not disorganized. Fragmented and possibly paranoid delusions are expected in the context of general response disorganization (Broen & Storms, 1966). However, it seems to be the case that a diagnosis of paranoid schizophrenia tends to be applied when relatively structured paranoid delusions are more prominent in the symptom picture and general response disorganization is somewhat less.

[3] Silverman (1964) has also written an excellent discussion of some research on attention in schizophrenia, using the concepts "extent of scanning" and "sensory input processing—ideational gating" which have some similarity to "breadth of observation." He also used a "field articulation" principle which approximates the inverse of "attention disorganization." A major purpose of our discussion is to relate the type of concepts emphasized by Silverman (1964) and Venables (1964) to the concept of "response disorganization" which seems useful in accounting for other research on schizophrenics (Broen & Storms, 1966).

[4] This statement does not contradict evidence that normals show consistent individual differences in scanning task stimuli (Gardner, 1961). The point is that, in addi-

Venables (1964) suggested that the reduced use of stimuli found in chronic schizophrenics is the result of increased arousal. This suggestion does not necessarily contradict our discussion in that there are several studies which suggest that increased arousal increases response disorganization in schizophrenics (Broen & Storms, 1966). Callaway and Stone (1960) have also suggested that arousal may lead to greater equivalence of probabilities of competing responses which people attempt to handle by reduced observation.

SUMMARY PROPOSITIONS

What has been implied thus far can be summarized as three propositions. Each proposition is based on processes which are hypothesized to affect major subgroups of the specified groups, and should be understood as referring to average group responding:

1. When a task has a single outstanding relevant stimulus which evokes several response tendencies, normals and schizophrenics will tend to have the same dominant and competing responses, but both acute and chronic schizophrenics will show more disorganized variation among these responses than normals. One area of empirical support for this proposition has been found when word-association tasks are repeated. Acute schizophrenics have greater variation in their associations than normals (Storms & Broen, 1964), and the same abnormal response instability has been found in poor premorbid chronics (Dokecki, Polidoro, & Cromwell, 1965).

2. In multiple stimulus tasks, acute schizophrenics will still be near normal in their breadth of observation and in the stimuli they observe, but their response disorganization means that attention hierarchies will be less organized than is the case for normals. Acute schizophrenics' attention will be abnormally scattered among the stimuli they have observed. Approximately normal breadth of observation together with disorganized attention means that acutes have attention scattered over a greater range of stimuli than normals or chronic schizophrenics.

3. In multiple stimulus tasks, chronic schizophrenics (especially nonparanoids) do not observe as many stimulus classes as normals or acute schizophrenics. The proportion of irrelevant to relevant stimuli observed by chronics depends on whether or not relevant and irrelevant stimuli are from different, easily discriminable categories, and on the relationship between stimulus pull and relevance. Like acutes, chronics have abnormal response disorganization so their attention will be abnormally variable, but only within the limited number of stimuli they have observed.

TWO KINDS OF OVERINCLUSION

We can now turn to the apparently contradictory statements. Statement *1a*, "Overinclusive concepts have been found in chronic schizophrenics," refers to the results of a task where single stimuli (words) were sorted into one of two conceptual classes (Chapman, 1961). From Proposition 1 above, if stimuli evoke both correct and incorrect sorting tendencies, chronics should make more sorting errors than normals, that is, should overinclude some words in incorrect classes. This is exactly what Statement 1*a* means. For example, in one part of the study, subjects were asked to sort the names of fruit, vegetables, and sports equipment into two classes: "fruit" and "not fruit." Sorting names of vegetables was most difficult. For normals the dominant response was to place vegetables correctly in the "not fruit" category, but presumably because of mediated generalization there was also a competing tendency to place vegetables in the fruit category. Like normals, chronic schizophrenics made most of their errors in sorting these objects which elicit competing responses. Chronic schizophrenics had the same dominant and competing re-

---

tion to these consistent individual differences, normals show temporary reductions in breadth of observation under increased perceptual load. Because attention disorganization (which increases perceptual load) is more constant and extreme in schizophrenics, chronics learn consistent and more extreme narrowing of observation.

sponses as normals, the difference was simply that the probabilities of dominant and competing responses were closer together for the chronics. In other words, chronics included more vegetables in the fruit category. This kind of overinclusion is simply disorganization among the alternate responses to the stimulus in single-stimulus tasks. As has been noted, acute schizophrenics also show disorganized responding in single stimulus-multiple response tasks, and similar results would be expected for acutes on this type of Chapman sorting task. Downing, Ebert, and Shubrooks (1963) showed that acutes tended to make the same kinds of errors as Chapman's chronics did on a grossly similar sorting task, but some task differences and the lack of results from a normal group make conclusions uncertain.

The meaning of overinclusion in Statement 1*b* is different. This statement refers to the number of different kinds of stimuli which are used in a multiple-stimulus task (Payne, 1962). Subjects were shown a number of objects and were asked to sort the objects repeatedly into groups that belong together, using as many different simple reasons for grouping the objects as they could think of. There were many possible stimulus categories which could be used. Normals tended to focus on stimuli such as shape of objects, color, or weight; only rarely using characteristics such as shape of shadows cast by the objects, or whether or not the objects were scratched. (With nondirective task requirements, attention was organized according to past learning about the utility of different stimulus classes.) For acute schizophrenics, normal observation together with not having as hierarchically organized attention should lead to responding to more stimuli than normals—using the stimuli that normals focus on, and also using the stimuli that tend to be ignored by normals (Proposition 2). However, because acutes are expected to use more stimuli than normals, boredom or fatigue effects must be considered. To the extent that these factors affect performance, acutes would not match normals in the use of the stimuli on which normals focus. The results (Payne, 1962) fit these expectations. Acutes produced relatively fewer sortings based on the kinds of stimuli normals tended to focus on, and more sortings based on stimuli that normals used only rarely (first part of Statement 1*b*). Overall, acutes used more types of stimuli than normals.

Chronic schizophrenics, with reduced range of observation, would have fewer kinds of stimuli eliciting competing focusing responses. Chronics have the basic schizophrenic response disorganization, but attention should be scattered only within the limited range of stimuli which are noted. In the Payne task, the stimuli which normals used most, and those stimuli normals used rarely, both included stimulus classes with relatively little stimulus pull. Reduced observation should, therefore, reduce noticing both kinds of stimuli (Proposition 3). Thus, one effect of reduced observation should be that chronics will show subnormal use of those stimuli on which normals usually focus. Payne (1962) found this reduced use of normally dominant stimuli in chronics. The picture should be different for stimuli that normals rarely focus on. For these stimuli, response disorganization and reduced observation should counteract each other in effect on differences between normals and chronics; reduced observation in chronics being offset by normals tending to focus on other stimuli. This is also consistent with Payne's data. Chronics did not use abnormally many unusual stimulus characteristics as bases for sorting. This is what is meant by lack of overinclusion in chronics in Statement 1*b*.

NARROW ATTENTION AND DISTRACTIBILITY IN CHRONIC SCHIZOPHRENICS

Recognizing that reduced range of observation and reduced organization of competing responses both occur in chronic schizophrenics may clarify the apparently contradictory statements that chronics have narrowed attention, and chronics are abnormally distractible

(Statements 2*a* and *b*). Research results depend on the research task. The extent of the chronic's distractibility, relative to normals, depends on the likelihood that his limited observation will include stimuli which are considered to be distractors.

When the stimuli with strongest pull are relevant to task requirements, and other relevant and irrelevant stimuli are weaker, chronics will observe the stronger (relevant) stimuli; their subnormal observation only reflected in ignoring peripheral stimuli. Attention will be scattered only among the relevant stimuli that are observed. In the sense of reduced use of peripheral stimuli, chronics will not be more distractible than normals. This does not mean that chronics will do better than normals even when relevant stimuli are few and all have strong pull. In such situations normal broad observation followed by appropriate attention focus should offset the chronic's reduced observation of irrelevant stimuli. The clearest predictions regarding differences between normals and chronic schizophrenics are: (*a*) chronic schizophrenics should do relatively better when relevant stimuli have stronger pull than when all relevant stimuli are peripheral (e.g., chronics should show less impairment in directed intentional learning than in incidental learning); and (*b*) when relevant stimuli have strong pull but additional relevant stimuli are peripheral, chronics should do poorer than normals (e.g., in distance-constancy experiments). These predictions are consistent with considerable evidence which has been discussed in detail by Venables (1964). In these studies, the chronic's observation tends to be abnormally restricted to the relevant stimuli with strong pull, and the basic difference from normals is that chronics ignore weaker relevant stimuli which normals, with their broad observation followed by appropriate focus of attention, do use. This is what is meant by Statement 2*b*. The chronic's tendency to focus on what is relevant, together with ignoring other stimuli that are used by normals, is what may seem inconsistent with saying that chronic schizophrenics are distractible.

Chronic schizophrenics will be more distractible when there is little relationship, or a negative relationship between stimulus pull and stimulus relevance. In such cases the ratio of observation of relevant to irrelevant stimuli will not be high, and because reduced observation follows lines of stimulus pull, this ratio will be unchanged or made worse by reduced observation. Because of task requirements, in normals the initial observation will be followed by attention hierarchies which favor relevant stimuli. Attention will be more disorganized in chronics, leading to distraction by the relatively high proportion of irrelevant stimuli among the stimuli observed. A situation with little relationship between stimulus pull and relevancy is illustrated by another card-sorting task used by Chapman (1956). On each trial normals and chronic schizophrenics were instructed to place a response card with one of three sorting cards on the basis of figures on the response and sorting cards having a concept in common or being identical. There were three figures on each card. Subjects were instructed to use only the upper right figure on the response card although all three figures on each of the sorting cards were relevant. Note that reduced observation would not favor relevant stimuli on the basis of type of stimulus, because relevant stimuli could be any type of figure. Confining observation to relevant stimuli on the basis of position was also unlikely because different stimuli were close together, relevant stimuli were not in central position, and relevant stimuli were in different positions in response cards than on sorting cards. Thus, reduced observation would do little to favor noticing relevant as compared to irrelevant stimuli, and both normals and chronic schizophrenics should observe irrelevant as well as relevant stimuli. However, the more organized response hierarchies in normals would do two things: (*a*) provide more attention focus on relevant stimuli; and (*b*) give clearer dominance of appropriate over competing sorting tendencies if both relevant and irrelevant commonalities are attended to on a single trial. So normals should, and did (Chapman, 1956), make fewer sortings based on irrelevant stimuli. This is the meaning of "chronic schizophrenics are more distractible than normals" in Statement *2b*.

REFERENCES

BROADBENT, D. E. *Perception and communication.* Oxford: Pergamon Press, 1958.

BROEN, W. E., JR., & STORMS, L. H. Lawful disorganization: The process underlying a schizophrenic syndrome. *Psychological Review*, 1966, 73, 265-279.

BURSILL, A. E. The restriction of peripheral vision during exposure to hot and humid conditions. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 113-129.

CALLAWAY, E., & STONE, G. Re-evaluating focus of attention. In L. Uhr & J. G. Miller (Eds.), *Drugs and Behavior.* New York: Wiley, 1960. Pp. 393-398.

CAMERON, N. Reasoning, regression, and communication in schizophrenia. *Psychological Monographs*, 1938, **50**(1, Whole No. 221).

CHAPMAN, L. J. The role of type of distractor in the "concrete" conceptual performance of schizophrenics. *Journal of Personality*, 1956, **25**, 130–141.

CHAPMAN, L. J. A reinterpretation of some pathological disturbances in conceptual breadth. *Journal of Abnormal and Social Psychology*, 1961, **62**, 514-519.

DOKECKI, P. R., POLIDORO, L. G., & CROMWELL, R. L. Commonality and stability of word association responses in good and poor premorbid schizophrenics. *Journal of Abnormal Psychology*, 1965, **70**, 312–316.

DOWNING, R. W., EBERT, J. N., & SHUBROOKS, S. J. Effects of three types of verbal distractors on thinking in acute schizophrenia. *Perceptual and Motor Skills*, 1963, **17**, 881-882.

FREEMAN, T. On the psychopathology of schizophrenia. *Journal of Mental Science*, 1960, **106**, 925-937.

GARDNER, R. W. Cognitive controls of attention deployment as determinants of visual illusions. *Journal of Abnormal and Social Psychology*, 1961, **62**, 120–127.

HARLOW, H. F. Learning set and error factor theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959. Pp. 492–537.

LESTER, J. R. Production of associative sequences in schizophrenia and chronic brain syndrome. *Journal of Abnormal and Social Psychology*, 1960, **60**, 225–233.

MAHER, B. A. *Principles of psychopathology.* New York: McGraw-Hill, 1966.

MCGHIE, A., & CHAPMAN, J. Disorders of attention and perception in early schizophrenia. *British Journal of Medical Psychology*, 1961, **34**, 103–116.

PAYNE, R. W. An object classification test as a measure of overinclusive thinking in schizophrenic patients. *British Journal of Social and Clinical Psychology*, 1962, **1**, 213-221.

PAYNE, R. W. The measurement and significance of overinclusive thinking and retardation in schizophrenic patients. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of schizophrenia.* New York: Grune & Stratton, 1966. Pp. 77–97.

SHAKOW, D. Segmental set: A theory of the formal psychological deficit in schizophrenia. *Archives of General Psychiatry*, 1962, **6**, 1–17.

SILVERMAN, J. The problem of attention in research and theory in schizophrenia. *Psychological Review*, 1964, **71**, 352–379.

STORMS, L. H., & BROEN, W. E., JR. Verbal associative stability and appropriateness in schizophrenics, neurotics, and normals as a function of time pressure. *American Psychologist*, 1964, **19**, 460. (abstract)

VENABLES, P. H. Input dysfunction in schizophrenia. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 1. New York: Academic Press, 1964. Pp. 1–47.

ZAHN, T. P., ROSENTHAL, D., & SHAKOW, D. Effects of irregular preparatory intervals on reaction time in schizophrenia. *Journal of Abnormal and Social Psychology*, 1963, **67**, 44–52.

(Early publication received April 22, 1966)

Psychological Review
1966, Vol. 73, No. 6, 586–588

ON THE INCOMPATIBILITY OF THE HOUSTON AND OSGOOD TRANSFER SURFACES

CHARLES P. THOMPSON

Kansas State University

A comparison of the Houston and Osgood transfer surfaces demonstrates that they are incompatible.

Osgood's (1949) transfer surface predicts the direction and magnitude of transfer as a function of the degree of S_1-S_2 and R_1-R_2 similarity. Recently, Houston (1964) has suggested a relabeling of this surface using the degree of S_1-R_2 and S_2-R_1 similarity for the coordinates. The two surfaces are shown in Figure 1. Since Houston restricts the application of his surface to formal similarity, only that portion of Osgood's surface applicable to formal similarity is shown; that is, the extreme right portion of Osgood's surface representing response antagonism has been omitted.

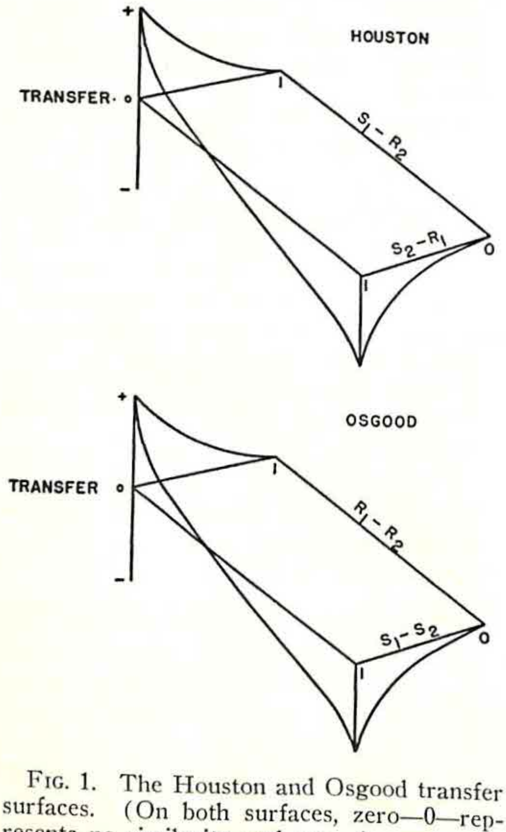

FIG. 1. The Houston and Osgood transfer surfaces. (On both surfaces, zero—0—represents no similarity and one—1—represents identity.)

These surfaces invite comparison, since both represent important attempts to summarize the effects of a potent variable in transfer. However, such comparison demonstrates that the two surfaces are incompatible.[1] This incompatibility is perhaps most strikingly evidenced in those instances in which some degree of similarity exists for at least one of the dimensions in both surfaces. Such instances may be produced by using three lists, two of which are totally dissimilar (hereafter designated as Lists A and B) and a third which is similar to both of the other lists (hereafter designated as List A/B). Sample lists of this type (using bigrams) are given in Table 1.

TABLE 1

SAMPLE BIGRAM LISTS OF THE A, B, AND A/B TYPE

List A	List B	List A/B
AM	WH	WA
RT	GH	GR
KE	FI	FE
XP	SC	SP
QU	CL	UL
AB	LD	AD
EX	IG	IE
PR	OW	PO
BU	NS	NU
TY	FO	OT

---

[1] It should be noted that the author assumes that both transfer surfaces are predicting performance relative to an A-B, C-D control paradigm. This seems reasonable since both surfaces indicate the A-B, C-D paradigm as a zero-transfer condition.

TABLE 2

TRANSFER PARADIGMS USING A, B, AND A/B LISTS AS STIMULI AND RESPONSES

13 transfer paradigms	Houston	Osgood
A-B A/B-B	–	+
A-B A/B-A	+	–
A/B-A B-A	0	+
A-A/B B-A	+	0
B-A A-A/B	?	0
B-A/B A-A/B	?	0
B-A B-A/B	0	?
B-A/B A/B-A	–(Max)	?
A/B-B A-A/B	0	?
A/B-B B-A	?	–
A/B-B A/B-A	?	–(Max)
B-A/B B-A	–	?
A-A/B A/B-A	+(Max)	?

Note.—Columns 2 and 3 indicate the direction of transfer predicted by Houston and Osgood.

Using A, B, and A/B lists as the stimuli and responses, one can construct a total of 13 functionally distinct transfer paradigms (not including trivial cases such as A-A) which meet the restriction of some degree of similarity obtaining for at least one of the dimensions in both transfer surfaces. These patterns are listed in Table 2. Of these, there are four conditions for which Houston and Osgood clearly make contradictory predictions. In two of these cases, one author predicts positive transfer while the other predicts negative transfer. The first of these two cases is the transfer paradigm A-B, A/B-B in which List A is used for the S_1's, List A/B is used for the S_2's, and List B is used for the R_1's and R_2's. In this case, Osgood predicts positive transfer since the responses are identical and the stimuli are similar. Houston, on the other hand, predicts negative transfer

TABLE 3

TRANSFER PARADIGMS USING A, B, A/B, AND C LISTS AS STIMULI AND RESPONSES

12 transfer paradigms	Houston	Osgood
C-A/B A-B	–	0
C-A A/B-A	–	0
C-A A-A/B	–(Max)	0
A-C A/B-A	0	–
A/B-C A-B	0	–
A-C A-A/B	0	–(Max)
A-A/B A-C	–	–(Max)
A/B-A A-C	–(Max)	–
A-B A/B-C	–	–
A-B C-A/B	0	0
A-A/B C-A	0	0
A/B-A C-A	0	0

Note.—Columns 2 and 3 indicate the direction of transfer predicted by Houston and Osgood.

since there is no similarity between S_1 and R_2 while S_2 and R_1 are similar. The second transfer paradigm for which opposite effects are predicted is the A-B, A/B-A condition. Here, since the responses are totally dissimilar and the stimuli are similar, Osgood predicts negative transfer. However, since S_1 and R_2 are identical and S_2 and R_1 are similar, Houston predicts positive transfer.

Indeed, an inspection of these 13 transfer paradigms demonstrates that, in all cases, Houston and Osgood predict different points on the transfer surface. However, since *(a)* a large portion of the transfer surface predicts no transfer, *(b)* assumptions about location must be made at points other than identity or total dissimilarity, and *(c)* the surfaces are obviously not of equal magnitude (e.g., consider the point of maximum transfer for both surfaces), Houston and Osgood do not necessarily make contradictory predictions in all such instances.

Adding another dissimilar list (hereafter designated List C) generates, under the restrictions noted above, 12 more functionally distinct transfer conditions. These paradigms are listed in Table 3. There are two points to be noted for these conditions: *(a)* The A-B, A/B-C and A-B, C-A/B paradigms are the only patterns under consideration on which Houston and Osgood agree regarding their location on the transfer surfaces; and *(b)* while Houston and Osgood both predict negative transfer for the A-A/B, A-C and A/B-A, A-C paradigms, they disagree as to which pattern will produce the most negative transfer.

While this note has been concerned with pointing out the incompatibility of these two transfer surfaces, it should also be noted that both surfaces have obvious defects. For example, Osgood predicts no transfer in the A-B, B-A paradigm (e.g., positive transfer found by Murdock, 1956) while Houston predicts that learning (i.e., A-B, A-B) will not take place. Nevertheless, the relationships suggested by the surfaces are important, and research investigating these relationships (and, hopefully, utilizing some of the paradigms suggested in this paper) should make an important contribution to our understanding of transfer.

REFERENCES

HOUSTON, J. P. Verbal transfer and interlist similarities. *Psychological Review,* 1964, **71**, 412–414.

MURDOCK, B. B., JR. "Backward" learning in paired associates. *Journal of Experimental Psychology,* 1956, **51**, 213–215.

OSGOOD, C. E. The similarity paradox in human learning: A resolution. *Psychological Review,* 1949, **56**, 132–143.

(Received September 27, 1965)